New reporters of protein trafficking and protein-protein interactions in live cells

by

Marta Fernández Suárez

B.S. Chemical Engineering (1999)
Institut Químic de Sarrià (Spain)

M.S. Chemical Engineering (2001)
Institut Químic de Sarrià (Spain)

Submitted to the Department of Chemistry
in Partial Fulfillment of the Requirements for the
Degree of Doctor of Philosophy

at the

Massachusetts Institute of Technology

June 2008

Signature of the Author: _____
Department of Chemistry
May 29, 2008

Certified by: _____
Alice Y. Ting
Pfizer-Laubach Associate Professor of Chemistry
Thesis Supervisor

Accepted by: _____
Robert W. Field
Chairman, Departmental Committee on Graduate Students

This doctoral thesis has been examined by a committee of the Department of Chemistry as follows:

Barbara Imperiali
Class of 1922 Professor of Chemistry and Professor of Biology

Alice Y. Ting
Pfizer-Laubach Associate Professor of Chemistry
Thesis Supervisor

David Bartel
Professor of Biology
Member, Whitehead Institute
Investigator, Howard Hughes Medical Institute

New reporters of protein trafficking and protein-protein interactions in live cells

by

Marta Fernández Suárez

Submitted to the Department of Chemistry
on May 29, 2008 in partial fulfillment of the
requirements for the Degree of Doctor of Philosophy

ABSTRACT

Here, we describe our attempts to harness the exquisite specificity of natural protein and RNA enzymes to develop improved methods to study protein localization and protein-protein interactions in live cells. We first attempted to detect endogenous protein-protein interactions (PPIs) in live cells by means of a ribozyme complementation assay, but we found that the strategy was limited by the interaction affinity constraints and by low ribozyme activity in cells. We then sought to still detect interactions among endogenous proteins but in fixed cells. We devised an improved immunofluorescence (IF) technique, in which the antibodies are conjugated to an enzyme-substrate pair. We chose *E. coli* biotin ligase (BirA), which catalyzes the covalent ligation of biotin to a 15-amino acid recognition sequence (AP). Only upon PPI would BirA be in close enough proximity to biotinylate the AP. Although the use of proximity biotinylation within the IF scheme proved challenging because of the geometric rigidity of the antibody conjugates, we later successfully applied the concept to the study of recombinant proteins in live cells, where BirA and AP were each genetically fused to the proteins of interest. We demonstrated that this method offers a combination of high spatial and temporal resolution with a low rate of false positives. We engineered the BirA/AP affinity to reduce background and eliminate false positives, while still allowing robust detection of relatively transient PPIs (half-life > 1 minute). We demonstrated that the methodology exhibits high specificity for the detection of PPIs in living mammalian cells, with a fold induction in the detected signal upon PPI of ~ 5–25. Using FRB-FKBP12 system as a model, the BirA/AP(-3) pair was also able to quantitatively predict interaction K_ds. Importantly, we showed that proximity biotinylation can detect the subcellular localization of the PPI under study.

We also developed a new method for site-specific labeling of proteins in live cells. Through rational design, we re-directed *E. coli* lipoic acid ligase (LplA) to specifically ligate an unnatural alkyl azide substrate to an engineered 22-amino acid LplA acceptor peptide (LAP) tag. The alkyl azide can then be selectively derivatized with a cyclooctyne conjugated to any probe of interest. We first demonstrated that LplA can be used to label LAP-tagged proteins with Cy3, AlexaFluor568, and biotin at the surface of living mammalian cells, and we then applied the methodology to one- and two-color cell-surface receptor labeling. Finally, we also showed that LplA can site-specifically label intracellular proteins, although the signal/background ratio still needs to be improved.

Thesis Supervisor: Alice Y. Ting
Title: Associate Professor of Chemistry

Acknowledgements

First, and foremost, I want to thank my advisor and mentor Prof. Alice Ting for her tremendous support throughout these five years. Her passion for science has been a continuous source of inspiration. Alice emanated optimism when I thought things would never work, and she offered her objective criticism at times when I became overexcited. From her, I learnt to dream big, while keeping the focus in the details. At the beginning, she taught me every technique in molecular biology; at the end, she trusted my judgment and encouraged me to lead my own experiments. I also want to thank her for offering me the opportunity to give several conference talks, which proved instrumental for my thesis defense and post-doc interview, and that will be of great help in my career as a scientist.

I also want to thank my former advisors Prof. Elazer Edelman and Dr. Mercedes Balcells. My one-year stay in Elazer's lab changed my career path. I was a Chemical Engineer, and the idea of pursuing a PhD had never crossed my mind. Elazer and Merche changed that course. Elazer enlightened me on the potential of an academic career to deeply impact the world, whereas Merche showed me that doing research was at the same time, challenging and fun. It was thanks to Merche that I came to MIT in the first place, and it is thanks to her that I finally decided to become a doctor.

I thank Prof. Barbara Imperiali for her words of courage at our annual meetings and Prof. David Bartel for his early advice on the ribozyme project. I thank them both for their helpful comments on this thesis. I thank Prof. JoAnne Stubbe for helpful discussions on enzyme folding and kinetics and for being my best teacher ever. I am equally grateful to Sally Susnowitz and Amy Smith. Working with them helped me balance my life in the lab with my passion for development. I thank the La Caixa Foundation and the Lewis Paul Chapin Fund for funding my graduate work.

I feel blessed with the opportunity to interact with so many great scientists in the Ting Lab. I want to thank Dr. Ed Curtis, Dr. Dan Chinnapen, Scott Chen, and Dr. Yi Zheng for their contributions. I am most grateful to everyone in the 'LplA team': Hemanta, Katie, and Laura. Working together was an amazing team-work experience, enhanced by their enormous individual talents, both as scientists and as friends. I also want to thank Sarah, Sujiet, Irwin, and Chi-Wang for their friendship and words of courage when they were most needed. I want to thank all the members of the Ting lab for making my time at MIT a fun and enriching experience. I also owe a big part of my success in graduate school to my friends outside MIT. Patricia, the 11-member Mellens family, Laure-Anne, and Stella helped me in balancing work with life.

I am deeply grateful to my parents. It is thanks to them that I am who I am. They have made a big effort to always give me the best and to support, economically and emotionally, all my endeavors. Both of them, a chemist and a doctor, injected me with passion for science, an immense curiosity for life, and the compelling need to use my talents for the benefit of others. I thank them because I know they make a big effort to support my choices even if that means being far away. And finally, I want to honestly and deeply thank Victor. Anything I can say here would not be enough. I thank him for his patience and understanding during the many days and weekends when I barely devoted time to him. I thank him for his enormous support during these last six months. I thank him for dragging me out of the lab at times and reminding me that there are other important things. I thank him for being here, and for being here his own way. I could have not made it so far without him, and I can just look forward to a future together.

Table of Contents

List of Figures

Part I : New reporters for the detection of protein-protein interactions

Chapter 1 : Introduction. Methodologies for detecting protein-protein interactions in cells

Chapter 2 : Development of an RNA-based detection of endogenous protein-protein interactions

List of Tables

List of Abbreviations

2OGDH	2-oxoglutarate dehydrogenase complex
ABC	ATP-binding cassette
ADH	Adipic dihydrazide
A_{260}	Absorbance at 260 nm
AP	Biotin ligase acceptor peptide
AP-2	Adaptor protein 2
ATP	Adenosine tri-phosphate
Az	Aptazyme (aptamer + ribozyme) fusion
Azide **7**-AMP	Azide **7** adenylate ester
AzL	Left monomer of a dimeric aptazyme
AzR	Right monomer of a dimeric aptazyme
BALK	Biotin-alkyne
BCA	Bicinchoninic acid
BiFC	Bimolecular fluorescence complementation
Biotin-AMP	Biotin adenylate ester
BirA	*E. coli* biotin ligase
Bla	β-lactamase
BMH	1,6-bismaleimido hexane
bp	Base pair
BRET	Bioluminescence resonance energy transfer
BSA	Bovine serum albumin
Cbl	Ubiquitin ligase
C_f	Final concentration
CFDA	Carboxyfluorescein diacetate
CFP	Cyan fluorescence protein
CHO	Chinese hamster ovary cells
CoA	Coenzyme A
CoCT	Co-transcriptional cleavage
CIP	Calf intestinal phosphatase
COS-7	African green monkey kidney cells
DCM	Dichloromethane
DEPC	Diethylpyrocarbonate
DHFR	Dihydrofolate reductase
Diamino-PEG	*O,O'*-Bis(3-aminopropyl)diethylene glycol
DIC	Differential interference contrast
DiD	1,1'-dioctadecyl-3,3,3',3'-tetramethylindodicarbocyanine
DMEM	Dulbecco's modified Eagle's medium
DMF	Dimethyl formamide
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DPBS	Dulbecco's phosphate-buffered saline
DPBS-B	DPBS + 3% BSA
DPDPB	1,4-di-[3'-(2'-pyridyldithio)-propionamido]butane

DTT	Dithiothreitol
E2p	9kD hybrid lipoyl domain derived from the second subunit of the *E. coli* PDH complex
E2p(ala)	Mutant of E2p with the lysine site of modification mutated to alanine
E. coli	*Escherichia coli*
EDA	ethylenediamine
EDC	1-ethyl-3-[3-dimethylaminopropyl]carbodiimide hydrochloride
EDTA	Ethylenediamine tetraacetic acid
EGF	Epidermal growth factor
EGFR	Epidermal growth factor receptor
eLplA	Same as LplA; "e" stands for *E. coli* gene
Eps15	Epidermal growth factor receptor pathway substrate 15
ERK	Extracellular signal-regulated kinase
ESI-MS	Electrospray ionization mass spectrometry
FBS	Fetal bovine serum
FKBP12	12 kD FK506 binding protein
FLIM	Fluorescence lifetime imaging microscopy
FLuc	Firefly luciferase
FP	Fluorescence protein
FRB	FKBP-rapamycin binding domain of the mTor protein
FRET	Fluorescence resonance energy transfer
Gab	GTPase-activating protein
GFP	Green fluorescence protein
gpTGase	Guinea pig transglutaminase
GlcN6P	Glucosamine-6-phosphate
Grb2	Growth factor receptor-bound protein 2
HA	Hemagglutinin
HCl	Hydrochloric acid
HDDA	*N,N'*-dimethyl-1,6-diaminohexane
HEK	Human embryonic kidney cells
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HHRz	Hammerhead ribozyme
hLplA	"Humanized" LplA (gene sequence optimized for mammalian expression)
HIV	Human immunodeficiency virus
hOAT	Human organic anion transporter
HP1	Histone protein 1
HPLC	High-performance liquid chromatography
IF	Immunofluorescence
IP	Immunoprecipitation
IPTG	Isopropyl-β-D-thiogalactopyranoside
IRS-1	Insulin receptor substrate-1
LAP	Lipoic acid ligase acceptor peptide
LAP(ala)	Mutant of LAP with the lysine site of modification mutated to alanine
LB	Luria-Bertani bacterial growth media
LDL	Low-density lipoprotein

LDLR	Low-density lipoprotein receptor
Lipoyl-AMP	Lipoic acid adenylate ester
LPL	Mammalian lipoate-protein ligase
LplA	*E. coli* lipoic acid ligase
LplA(arg)	Catalytically inactive mutant of LplA with Lys133 mutated to Arg
MALDI	Matrix-assisted laser desorption/ionization time of flight
MAPK	Mitogen-activated protein kinase
MEK	MAPK kinase or ERK kinase
MES	2-(N-morpholino)ethanesulfonic acid
$MgCl_2$	Magnesium chloride
MPBH	4-(4-*N*-maleimidophenyl) butyric acid hydrazide
MS	Mass-spectrometry
Mz	Maxizyme
MzL	Left monomer of a dimeric maxizyme
MzR	Right monomer of a dimeric maxizyme
NaCl	Sodium chloride
NAD^+	Nicotinamide adenine dinucleotide
NEM	*N*-ethylmaleimide
NFκB	Nuclear factor κB
NHS	*N*-hydroxysuccinimide
NLS	Nuclear localization signal
NMR	Nuclear magnetic resonance
NOESY	Nuclear Overhauser effect spectroscopy
nt	Nucleotide
NTA	Nitrilotriacetic acid
OCT	Monofluorinated cyclooctyne
OD	Optical density
ORF	Open reading frame
PAGE	Polyacrylamide gel electrophoresis
PEG	Polyethyleneglycol
PBS	Phosphate-buffered saline
PBS-Mg	PBS + 5 mM magnesium chloride
PBS-T	PBS + 0.1% Tween 20
PCA	Protein complementation assays
PCR	Polymerase chain reaction
PDH	Pyruvate dehydrogenase complex
pEGFR	Phosphorylated epidermal growth factor receptor
PET	Positron emission tomography
PFA	Paraformaldehyde
PFP	Pentafluorophenyl
PKB	Protein kinase B
PKC	Protein kinase C
PLCγ	Phospholipase C γ
PMSF	Phenylmethylsulphonyl fluoride
PAGE	Polyacrylamide gel electrophoresis
PI3K	Phosphatidyl inositol 3-kinase

PPi	Pyrophosphate
PPI	Protein-protein interaction
PPTase	Phosphopantetheine transferase
PTB	Phosphotyrosine binding motif
RHR	Rel homology region
RIPA	Radioimmuno precipitation assay
RLuc	*Renilla* luciferase
RNA	Ribonucleic acid
RNase	Ribonuclease
rNTP	Ribonucleotide tri-phosphate
s.d.	Standard deviation
Rz	Ribozyme
SDS	Sodium dodecyl sulfate
SELEX	Systematic evolution of ligands by exponential enrichment
SH2	Src homology 2
Shc	Src homologus and collagen-like protein
Shp1	Src homology phosphatase 1
SMCC	Succinimidyl 4-[*N*-maleimidomethyl]cyclohexane-1-carboxylate
STORM	Stochastic optical reconstruction microscopy
Sos	Son of sevenless protein
Ta	*Thermoplasma acidophilum*
TAP	Tandem affinity purification
TBS	Tris-buffered saline
TBS-T	TBS + 0.05% Tween 20
TCA	Trichloroacetic acid
TCEP	Tris(2-carboxyethyl)phosphine
TEA	Triethylamine
Tet	Tetracycline
TFA	Trifluoroacetic acid
TLC	Thin layer chromatography
TPP	Thiamine pyrophosphate
Tris	Trishydroxymethylaminomethane
WB	Western blotting
WT	Wild-type
YFP	Yellow fluorescence protein

Part I : New reporters for the detection of protein-protein interactions

Chapter 1 : Introduction. Methodologies for detecting protein-protein interactions in cells

Introduction

The recent characterization of several eukaryotic protein interaction networks has made it clear that protein interactions determine the outcome of most cellular processes[1-3]. Inside cells, protein-protein interactions are instrumental in enzymatic actions upon protein substrates and in the protein assemblies that govern signal transduction, cell division, DNA replication, and transcription initiation. Outside cells, protein interactions allow cells to talk with each other: ligands expressed on the cell surface often bind protein receptors expressed on adjacent cells, and secreted protein ligands bind receptors on distant cells. Therefore, identifying and characterizing protein–protein interactions (PPIs) and their networks is essential for understanding the mechanisms of biological processes on a molecular level.

Traditionally, the interactions between two recombinantly expressed and purified proteins were studied in a test tube, where PPIs are mainly governed by the interaction affinity between the protein partners. In contrast, in the cellular context, other factors such as spatial confinement, post-translational modifications, and hindering of interaction surfaces within protein complexes greatly affect the ability of any two given proteins to interact; ultimately generating a network of PPIs that is tightly regulated in space and time. For this reason, many new techniques have been developed in recent years that attempt to study protein-protein interactions in the cellular context.

This chapter describes the most common methods currently available to detect protein-protein interactions in the cellular context, that is, in cell extracts, in fixed cells, or in living cells and animals. Focus will be on describing how the methods work and what their weaknesses and strengths are, in order to set the stage for Chapters 2–4, where new approaches to overcome current limitations are proposed.

Detecting endogenous protein-protein interactions in cell extracts

The study of protein–protein interactions involving endogenous proteins frequently relies on the immunoaffinity capture of a protein of interest followed by identification of co-purifying interactors[4]. In this method, the so-called co-

immunoprecipitation technique (co-IP), the protein of interest is recognized and bound by a target antibody added to the cell extract. A second antibody, often bound at one end to a solid matrix, such as a bead, binds to the first antibody, precipitating the complex out of solution. If the antigen is associated in a complex of proteins, other members of the complex may be co-precipitated and hence separated from the rest of the cellular milieu. The subsequent identification of the co-precipitated interacting partners was originally performed by Western blotting (WB), but this approach required an *a priori* guess of the partner identity. Mass spectrometry is now increasingly being applied to the identification of unknown co-purifying interactors.

Although co-IP followed by WB is still regarded as the gold standard for the study of cellular PPIs, it has many problems. On the one hand, co-IP is characterized by a high rate of false positives. The biggest challenge on performing co-IP lies in distinguishing physiological interactors from unspecific binders; and how well this separation works depends on the affinity of the antibody for the antigenic site and on the concentration of the relevant proteins in the extract[5]. The problem of availability and affinity of a specific antibody can be solved by fusing an affinity tag to the protein of interest, but this method defeats the purpose of studying endogenous proteins. On the other hand, co-IP is also likely to miss transient interactions or interactions with low affinities, which dissociate during extract preparation, thus leading to false negatives[4]. Finally, because co-IP precipitates proteins from a cell extract, it does not provide any spatial or temporal information regarding the PPI.

Nonetheless, co-IP coupled to mass spectrometric (MS) analysis has become an unmatched tool for the identification of new PPIs[6]. Using MS analysis post-IP, quantitative proteomics can be used to identify false-positive interactions[7]. Quantitative proteomics is the common name for a set of techniques to differentially label proteins with stable isotopes and compare their relative abundance in samples analyzed side-by-side using MS. Using a combination of IP and MS, large-scale mapping of human PPIs has already been achieved[8].

Mass spectrometry has also been combined recently with the classic methods of affinity purification or tandem affinity purification[9] to yield enormous data sets and partially map the PPI networks of *Saccharomyces cerevisiae*[2,10,11].

Crosslinking is another classical approach for determining PPIs[12]. Crosslinking is commonly used as a strategy to minimize the presence of unspecific interactors during co-IP experiments. In this context, crosslinking is used to covalently link the constituents of protein complexes so that they can be subjected to stringent salt and detergent washes after co-IP. Crosslinking also offers the advantage of trapping transient interactions, which could otherwise not be detected by co-IP. Although non-specific crosslinking using for example paraformaldehyde has been successfully used[13,14], site-specific incorporation of a crosslinking probe (usually called photoaffinity probe) is usually preferred to minimize false positives. The best current methodology for site-specific photoaffinity probe incorporation is unnatural amino acid mutagenesis[15-17], but its application is still limited in mammalian cells because of the prevalence of amber stop codons and the low suppression efficiencies, which give rise to truncated by-products[18]. Other methods for site-specific incorporation of photoaffinity probes have been described, but their applicability in cells still needs to be demonstrated. For example, Liu et al. used the tetracysteine labeling tag (see Chapter 5 for a description) to site-specifically introduce a dihydrophenylalanine crosslinking probe onto the acidic domain of the VP16 transactivator to demonstrate its interaction with the 26S proteosome *in vitro*[19]. Better site-specific targeting methods need to be developed to fully realize the potential of the *in vivo* crosslinking technique.

Detecting endogenous protein-protein interactions in fixed samples

As a step forward from co-IP, the immunofluorescence (IF) technique provides information on the spatial localization of the interacting partners. However, because it requires cell fixation, it still lacks temporal resolution. In IF, fluorophore-conjugated antibodies against the target proteins are used to stain a fixed sample, and the presence of overlap between the two colors reports on protein co-localization[20]. Although IF has been very useful in studying the spatial distribution of proteins that are known interacting partners, it must be noted that IF cannot in reality distinguish whether two proteins are truly interacting or are just within the same focal spot in the microscope, that is, co-localized.

In conventional lens-based microscopy, such as wide-field or confocal microscopy, the spatial resolution of IF is limited by the diffraction of light[21]. This means that, using excitation light of ~ 400 nm wavelength, which is typical in cell microscopy, the maximum theoretical lateral spatial resolution that can be achieved is ~ 150 nm (or ~ 250 nm in practice)[22]. That is, with a conventional microscope, one cannot distinguish two features that are closer than 250 nm. Because protein sizes range from 25 Å x 50 Å for the green fluorescence protein (GFP) to ~ 250–500 nm for most antibodies, it follows that the resolution of IF is not enough to distinguish whether two proteins are merely co-localized or whether they are truly interacting. Many research groups are currently working on improving the spatial resolution of far-field microscopy by modifying the physics of fluorophore detection (for a review, see Hell[23]).

As a response to the need for improving the spatial resolution of IF, other groups have applied fluorescence resonance energy transfer (FRET) microscopy to the IF technique. FRET is a quantum mechanical process in which radiationless energy (an exciton) is transferred over a distance, usually < 80–100 Å, from one fluorophore to another by long-range dipole-dipole coupling[24]. Upon excitation of the donor fluorophore, the exciton is transferred and excites the acceptor fluorophore, which subsequently emits light that is further red-shifted than the donor emission. FRET requires that the emission spectrum of the donor overlaps with the excitation spectrum of the acceptor. The efficiency of the transfer depends on the spectral overlap and on the distance and relative orientation between the two fluorophores. In IF-FRET, each antibody is then conjugated to the donor or acceptor fluorophore. Because of the FRET dependence on the distance between fluorophores, the combination of FRET and IF results in the detection of proteins that are separated by distances of 1–10 nm[25]. Common fluorophore pairs used in IF-FRET are the cyanine dyes Cy3-Cy5, the Alexa Fluor 488-Alexa Fluor 555 dyes, and Fluorescein-Rhodamine. Unfortunately, FRET efficiency is inversely proportional to the sixth power of the separation between the donor and acceptor and it is highly dependent on the relative orientation of the fluorophores. This means that the IF-FRET technique requires that the binding of the antibodies against the target proteins places the two fluorophores in the appropriate orientation and within a

short distance range. This has proven difficult in many cases, and thus IF-FRET has not found widespread use for the detection of cellular PPIs.

Detecting protein-protein interactions in living cells and animals

In order to study PPIs with spatial and temporal resolution and within their native cellular context, many groups have tried in recent years to develop methods that can detect PPIs in living cells.

Yeast two-hybrid systems

The introduction of two-hybrid systems established a novel tool for high-throughput screening and discovery of new protein interactions by means of transcriptional activation of a reporter[26,27]. It has been estimated that more than 50% of all protein interactions described in the literature have been detected using the yeast two-hybrid system[28]. In this strategy, a transcription factor is dissected into two domains, a DNA-binding domain –required for transcription factor binding to DNA– and an activator domain – required for basal activation of transcription. Each protein of a potentially interacting partner is then fused to one of the split fragments of the transcription factor, such that interaction between the two proteins brings the two domains into close proximity. In such proximity, the two transcription factor fragments can re-assemble to result in transcription activation.

Although yeast two-hybrid systems revolutionized the study of PPIs, the strategy has the disadvantage of being limited temporally (i.e., it takes several hours from activation of transcription until enough protein is expressed and a detectable signal is generated). Additionally, two-hybrid systems require nuclear translocation of the hybrid fusion proteins, which is not always possible (e.g., for membrane proteins)[29]. To overcome these limitations, several types of reporters have been developed that utilize fluorescence and bioluminescence resonance energy transfer (FRET and BRET), or protein fragment complementation.

Fluorescence and bioluminescence resonance energy transfer

FRET has been applied to the study of PPIs in two manners. The simplest configuration is an intermolecular design in which each interacting protein is fused to a fluorescence protein (FP) donor (e.g., cyan fluorescence protein, CFP) or acceptor (e.g., yellow fluorescence protein, YFP)[30,31]. When the two proteins interact, the donor and the acceptor get close enough to each other to transfer the exciton, resulting in FRET. The most common FP pair used in intermolecular FRET studies is CFP-YFP, but other pairs have also been employed, such as BFP-GFP and CFP-dsRed. One of the main disadvantages of this design is the high rate of false negatives because the PPI may not result in FRET if the distance between the fluorophore pairs upon PPI is too large. In addition, false positives may also be seen if the fluorescence protein pairs come into close proximity even in the absence of a PPI owing to their intrinsic dimerization affinity. The latter effect can be minimized by using FPs with the A206K mutation, which eliminates dimerization[32]. Additionally, problems arise when the donor and acceptor fusion proteins are not present in an exact 1:1 ratio, which alters the FRET measurements; although this artifact can be avoided if fluorescence lifetime imaging microscopy (FLIM) or emission anisotropy are used for the recording manifold[24,33].

The second FRET design is based on an intramolecular interaction[34]. In this configuration, the two interacting pairs as well as the FRET donor and acceptor are on a single polypeptide chain. The intramolecular design ensures that donor and acceptor are present in a 1:1 ratio and it also eliminates false negatives. Unfortunately, the rate of false positives increases in this design because of the increased effective concentration of the detector pair as well as the interacting proteins. Although there are some examples of intramolecular FRET applications to PPIs[34], this design has been predominantly used to study enzymatic activity (e.g., protease, kinase, and phosphatase activity), membrane potential, and Ca^{2+} oscillations[35].

For imaging in animals, bioluminescence is often preferred over fluorescence because it eliminates the need for an excitation beam, which can result in tissue damage. Additionally, bioluminescence resonance energy transfer (BRET) eliminates background cell autofluorescence caused by the excitation light, avoids background from donor emission, and exhibits an extremely high quantum yield, which results in more sensitive

measurements of cellular interactions. BRET works much like FRET except that the donor molecule is not a fluorescence protein or dye, but rather a bioluminescence protein[36]. Currently used BRET pairs consist of *Renilla* luciferase coupled with either modified EYFP or GFP[37,38]. In this design, *Renilla* luciferase catalyzes the oxidation of its substrate coelenterazine to produce light with a λ_{max} of 395 nm. This emission excites a modified GFP, which then emits the transferred energy at 508 nm. The main disadvantage of BRET is that no pair has been reported to date that is sufficiently red-shifted for optimal use *in vivo*[39].

Protein fragment complementation assays

Protein complementation assays (PCAs) represent an evolution of the yeast two-hybrid systems, in which the transcription factor is substituted by a reporter protein or enzyme. In PCAs, a monomeric reporter protein is rationally divided into two separate inactive components that can reconstitute function upon association. When these reporter fragments are fused to interacting proteins, the reporter is re-activated upon association of the interacting proteins[40]. PCAs are superior to the traditional two-hybrid methods for PPI detection, because they can be used in any cell type and within all subcellular compartments. Because PCA strategies take advantage of the all-or-none nature of protein folding, the signal generated has a large dynamic range over very narrow conditions, unlike FRET measurements. PCAs are also more sensitive and technically less challenging to implement than FRET. However, existing PCAs are limited either by their spatial resolution (e.g., split-β-lactamase and split-luciferase) or their temporal resolution (e.g., split-GFP).

The first PCA developed was based on the enzyme dihydrofolate reductase (DHFR)[41]. DHFR converts dihydrofolic acid into tetrahydrofolic acid, which in turn is transformed into many cellular cofactors. The DHFR-based PCA can be used in two different ways. The first one is a survival assay, in which null cells lacking DHFR can only survive if the two target proteins interact and reconstitute DHFR activity[42]. The second assay uses a fluorogenic derivative of methotrexate (i.e., an inhibitor of DHFR), which binds intact DHFR[43]. While the first method suffers from lack of temporal

information, the second one poses great toxicity to the cells under study because methotrexate inhibits the metabolism of folic acid. For these reasons, the split-DHFR has not been applied to any biologically relevant studies.

Split-β-galactosidase was developed next and offered some interesting advantages over split-DHFR such as signal amplification, mammalian live-cell compatibility, and the fact that its activity can be quantified by biochemical, fluorescent, or chemiluminescence readouts[44,45]. The main weakness of split-β-galactosidase is that it requires substrate treatment ex-*vivo*, hence losing any temporal information on the PPI under study.

Split-GFP was the first PCA that was able to provide great spatial resolution combined with some temporal information. Additionally, a unique characteristic of split-GFP systems is that they do not require an external small-molecule substrate. The first demonstration of split-GFP was based on a protein splicing system[46]. Protein splicing is a naturally occurring posttranslational processing event in which an intervening intein domain excises out of a precursor polypeptide in an autocatalytic fashion with concomitant linkage of the two flanking extein sequences by a native peptide bond[47]. In the first report of split-GFP, Ozawa and co-workers fused the N- and C-terminal halves of a yeast intein to the N- and C-terminal halves of GFP, and each of these fusion proteins was in turn linked to one of the interacting partners[46]. When interaction occurred between the two proteins, the N- and C- terminal halves of the intein were brought into close proximity and underwent correct folding, which induced a splicing event, thus linking the N- and C-terminal fragments of GFP directly by a peptide bond. The matured GFP formed the fluorophore, and the extent of the protein-protein interaction was evaluated by measuring the fluorescence intensity. Later, this and other split-FP fragments were also used in a configuration similar to the yeast two-hybrid without the use of an intein. The split-FP assay, also called bimolecular fluorescence complementation (BiFC) has been used to study the interactions between the transcriptions factors Jun-Fos[48] and Myc-Max-Mad[49] with high spatial resolution, allowing for subcellular localization. Although the temporal resolution of GFP has traditionally been hampered by the need for GFP maturation, which can take up to several hours, recent work by Demidov and co-workers reported a split-GFP that re-combined and matured in just a few minutes upon DNA hybridization[50]. It remains to be

seen whether this maturation kinetics still hold when the two GFP fragments are fused to proteins and expressed in living cells.

PCAs that employ enzymes instead of proteins offer the added advantage of signal amplification. The most widely used enzyme complementation assays are those based on β-lactamase and *Renilla* or firefly luciferase. Split-β-lactamase and split-luciferase are superior to the previously described split-DHFR and split-β-galactosidase because their fragments are smaller (~ 20 kD), there is no endogenous β-lactamase or luciferase acivity in mammalian cells, and they both use highly sensitive cell-permeant substrates as a readout. On top of signal amplification, these two systems exhibit immediate temporal response to a PPI thus improving the temporal resolution of split-GFP, but their reliance on diffusible small molecule substrates means that they cannot provide subcellular localization.

β-lactamase was first used as a reporter system by the Tsien laboratory (U.C. San Diego)[51]. Zlokarnik and co-workers designed a β-lactamase substrate in which two fluorophores are attached to the 7 and 3′ positions of a cephalosporin ring. In this configuration, the two fluorophores are close enough to exhibit efficient fluorescence resonance energy transfer. However, β-lactamase attack splits off the fluorophore on the 3′ position, disrupts FRET, and reestablishes fluorescence emission by the coumarin donor, thus resulting in an increase in the cellular blue-to-green emission signal ratio. Although β-lactamase has been successfully used to monitor the self-[52] and trans-splicing[53] activity of the Tetrahymena group I intron ribozyme, its use to study PPIs has been restricted to a few proof-of-principle demonstrations. For example, Wehrman and co-workers, the first ones to use split-β-lactamase, demonstrated its use on the known interactions between the Jun and Fos transcription factors and between FK206 binding protein (FKBP) and the FKBP-rapamycin binding domain of the mTor protein (FRB)[54]. In parallel, a different group also developed the same split-β-lactamase PCA and applied it to the proof-of-principle interactions between the apoptotic proteins Bcl2 and Bad, the homodimerization of Smad3, and the rapamycin-induced interaction between FKBP and FRB[55]. One of the main reasons why β-lactamase has not been used more widely is that its fluorescence substrate is not taken up equally by different cell types, and, in many cases, not at all (e.g., plant and yeast)[56].

Split-luciferase offers the same advantages over β-lactamase that we already mentioned for BRET, based on the fact that luciferase employs a bioluminescence readout. As previously mentioned, luciferase enzymes oxidize their substrate with concomitant emission of light. The two luciferase reporters most commonly used in optical imaging are *Renilla* luciferase, which uses coelenterazine as a substrate, and firefly luciferase, which oxidizes D-luciferin. *Renilla* luciferases (RLuc) generally emit blue, a property less favorable for *in vivo* imaging, whereas firefly luciferases (FLuc) generally emit yellow to red, enhancing their utility *in vivo*[39]. Additionally, several coelenterazine susbtrates of RLuc have been found to be substrates of endogenous mammalian proteins, which raises concerns on the invasiveness of the strategy[57]. As with GFP, luciferases were first used to detect PPIs in an intein-mediated split-firefly luciferase design[58]. Ozawa and co-workers used this technique to study insulin-mediated phosphorylation of the insulin receptor substrate-1 (IRS-1) and its target p85 subunit of phosphatidylinositol 3-kinase (PI3K). Although insulin is reported to trigger the kinase function of the insulin receptor within minutes, the bioluminescence signal was not observed until after 3 hours of insulin stimulation, which suggests that the split-intein system suffers from slow kinetic response rates[39]. In addition, high background luminescence was observed in cells, possibly owing to the splicing event occurring even when there was partial association of the intein fragments. Thereafter, both RLuc and FLuc were used in split-luciferase complementation assays. One of the most successful examples of split-RLuc is that of Kaihara and co-workers, which showed 25-fold signal induction after 5 minutes of insulin stimulation when they studied the insulin-mediated interaction between IRS-1 and PI3K[59]. In recent years, several groups have tried to improve the kinetics of split-FLuc; Paulmurugan and co-workers being the first ones to demonstrate PCAs in living small animals using the interaction of MyoD and Id as test proteins[60]. Probably the most impressive system was developed in the laboratory of Piwnica-Worms, where split-FLuc resulted in 23-fold signal induction upon interaction of FKBP and FRB in living mice[61]. The versatility of this reporter was also demonstrated by applying it to the study of two other interactions, that of Cdc25C and 14-3-3ε and the homodimerization of the STAT-1 transcription factor[61]. Recently, a major step forward has been taken recently by Stefan and co-workers, with the development of a reversible

split-RLuc, which allowed them to study the dynamics of G-protein coupled receptor-induced assembly and disassembly of protein kinase A[62].

Finally, although most PCAs have only been demonstrated for proof-of-principle examples, such as the known interaction between FKBP and FRB, some of them have already been capable of discovering new interactions and mapping out biochemical pathways. For example, Michnick and co-workers used split-DHFR to map a signal transduction network that controls initiation of translation in eukaryotes, and identified 14 interactions, of which five have not been observed previously[63]. Most impressively, the Michnick group has just recently used the same split-DHFR system to perform a genomewide *in vivo* screen for PPIs in *Saccharomyces cerevisiae*, where they identified 2,770 interactions, most of which had not been previously reported[64]. Using split-GFP, Kerppola and co-workers compared the interaction affinity of the different members of the Myc/Max/Mad families of transcription factor, and found that, contrary to what had been previously believed, Mad4 is a better interacting partner and can compete with Myc for binding to Max[49]. Split-GFP was also used in a cDNA library screening, which identified novel substrates and regulators of the serine/threonine protein kinase B (PKB)[65]. Thereafter, using a combination of split-β-lactamase and split-YFP, Remy and co-workers discovered a new interaction between PKB and the transcription factor Smad3, which is mediated by insulin but inhibited by transforming growth factor β (TGF-β), thus finding a link between the TGF-β and the growth factor signaling cascades[66].

Conclusions

Protein-protein interactions are involved in every cellular process ranging from gene expression and signal transduction to cell division and differentiation. Within this context, the goal of any method that aims to study protein-protein interactions would be to reveal not only the partners of particular proteins, but also how tight the interactions are, which surfaces they use to contact, and where and when contacts occur. Standard biochemical methods have yielded most of the available information about such

interactions, but these assays are often limited by the available reagents (e.g., monoclonal antibodies) or lack the appropriate cellular context. The development of fusion-protein assays such as the yeast two-hybrid method and the protein-complementation assays has expanded the potential for studying protein-protein interactions in intact, and specifically mammalian, cells greatly. Table 1-1 summarizes the weaknesses and strengths of the methods described in this chapter. It can be readily concluded that none of the existing methods has yet reached perfection. Furthermore, a comparative study has recently demonstrated that the combination of the high false-positive and false-negative rates of most of the available methods results in the fact that only a small percentage of yeast PPIs is supported by more than one method[67].

Ideally, a method for detecting protein-protein interactions should detect endogenous proteins in a non-invasive manner, with high temporal and spatial resolution, and with sufficient sensitivity to allow single cell imaging. The following Chapters 2–4 attempt to bridge some of these characteristics. Chapter 2 describes a concept to detect endogenous protein-protein interactions in live cells. This method represents the ultimate dream on PPIs detection. Unfortunately, the concept proved difficult to translate into a real and robust system. For this reason, we sought to develop less ambitious but yet powerful techniques. Chapter 3 describes an attempt to improve the spatial resolution of the commonly used immunofluorescence technique so that it can report on real PPIs rather than mere protein co-localization. Like the classic IF technique, this method can detect endogenous PPIs but it requires cell fixation and therefore lacks temporal information. In contrast, Chapter 4 describes a complementary method, which works in live cells, thus preserving spatial information, but detects recombinant fusion proteins. The method relies on the enzymatic transfer of biotin by the enzyme biotin ligase to its acceptor peptide and it represents an improvement over previously reported protein fragment complementation assays owing to its combined high sensitivity, improved spatial resolution, and ability to detect transient protein-protein interactions.

Strategy	Readout	Lysate	Fixed cells	Live cells	Living animals	Endogenous proteins	New PPIs [1]	Spatial resolution	Temporal resolution
Co-IP/WB	WB	√	–	–	–	√	–/√	–	–
Co-IP or TAP/MS	MS	√	–	–	–	√	√	–	–
IF	Fluorescence	–	√	–	–	√	–/√	–/√ [2]	–
IF-FRET	Fluorescence	–	√	–	–	√	–/√	√	–
FRET	Fluorescence	√	–	√	–	–	–/√	√	√
BRET	Bioluminescence	√	–	√	√	–	–/√	–/√ [3]	√
Two-hybrid	Fluorescence, bioluminescence, PET	–	–	√	√	–	–/√	–	–/√ [4]
Split-DHFR	Survival assay, fluorescence	√	–	√	–	–	–/√	–/√ [3]	–/√ [5]
Split-β-galactosidase	Fluorescence, absorbance	√	–	√	–	–	–/√	–	–
Split-GFP	Fluorescence	√	–	√	–	–	–/√	√	–/√ [6]
Split-β-lactamase	Fluorescence	√	–	√	–	–	–/√	–/√ [3]	√
Split-luciferase	Bioluminescence	√	–	√	√	–	–/√	–/√ [3]	√

Table 1-1: Strategies for detecting protein-protein interactions. Weaknesses and strengths. PPI: protein-protein interaction. IP: Immunoprecipitation. WB: Western blotting. TAP: tandem-affinity purification. MS: mass spectrometry. IF: immunofluorescence. FRET: fluorescence resonance energy transfer. BRET: bioluminescence resonance energy transfer. DHFR: dihydrofolate reductase. GFP: green fluorescence protein. PET: positron emission tomography. (1) The symbol –/√ means that the methods allows confirmation of a new PPI but it requires an *a priori* guess of such an interacting partner. (2) As explained in the text, IF can only demonstrate co-localization but not true PPI. (3) The small-molecule substrates used in these strategies are diffusible, thus loosing subcellular localization. (4) Requirement for nuclear translocation and transcription activation delays the signal by several hours. (5) The survival assay does not have temporal resolution. (6) Requirement for GFP maturation may delay the signal by up to few hours.

References

1. Giot,L. *et al.* A protein interaction map of Drosophila melanogaster. *Science* **302,** 1727-1736 (2003).

2. Krogan,N.J. *et al.* Global landscape of protein complexes in the yeast Saccharomyces cerevisiae. *Nature* **440**, 637-643 (2006).

3. Li,S. *et al.* A map of the interactome network of the metazoan C. elegans. *Science* **303**, 540-543 (2004).

4. Phizicky,E.M. & Fields,S. Protein-protein interactions: methods for detection and analysis. *Microbiol. Rev.* **59**, 94-123 (1995).

5. Markham,K., Bai,Y. & Schmitt-Ulms,G. Co-immunoprecipitations revisited: an update on experimental concepts and their implementation for sensitive interactome investigations of endogenous proteins. *Anal. Bioanal. Chem.* **389**, 461-473 (2007).

6. Kocher,T. & Superti-Furga,G. Mass spectrometry-based functional proteomics: from molecular machines to protein networks. *Nat. Methods* **4**, 807-815 (2007).

7. Ranish,J.A. *et al.* The study of macromolecular complexes by quantitative proteomics. *Nat. Genet.* **33**, 349-355 (2003).

8. Ewing,R.M. *et al.* Large-scale mapping of human protein-protein interactions by mass spectrometry. *Mol. Syst. Biol.* **3**, 89 (2007).

9. Berggard,T., Linse,S. & James,P. Methods for the detection and analysis of protein-protein interactions. *Proteomics.* **7**, 2833-2842 (2007).

10. Gavin,A.C. *et al.* Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* **415**, 141-147 (2002).

11. Ho,Y. *et al.* Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry. *Nature* **415**, 180-183 (2002).

12. Agou,F., Ye,F. & Veron,M. In vivo protein cross-linking. *Methods Mol. Biol.* **261**, 427-442 (2004).

13. Schmitt-Ulms,G. *et al.* Binding of neural cell adhesion molecules (N-CAMs) to the cellular prion protein. *J. Mol. Biol.* **314**, 1209-1225 (2001).

14. Vasilescu,J., Guo,X. & Kast,J. Identification of protein-protein interactions using in vivo cross-linking and mass spectrometry. *Proteomics.* **4**, 3845-3854 (2004).

15. Chin,J.W. & Schultz,P.G. In vivo photocrosslinking with unnatural amino Acid mutagenesis. *Chembiochem.* **3**, 1135-1137 (2002).

16. Farrell,I.S., Toroney,R., Hazen,J.L., Mehl,R.A. & Chin,J.W. Photo-cross-linking interacting proteins with a genetically encoded benzophenone. *Nat. Methods* **2**, 377-384 (2005).

17. Hino,N. *et al.* Protein photo-cross-linking in mammalian cells by site-specific incorporation of a photoreactive amino acid. *Nat. Methods* **2**, 201-206 (2005).

18. Liu,W., Brock,A., Chen,S., Chen,S. & Schultz,P.G. Genetic incorporation of unnatural amino acids into proteins in mammalian cells. *Nat. Methods* **4**, 239-244 (2007).

19. Liu,B., Archer,C.T., Burdine,L., Gillette,T.G. & Kodadek,T. Label transfer chemistry for the characterization of protein-protein interactions. *J. Am. Chem. Soc.* **129**, 12348-12349 (2007).

20. Miyashita,T. Confocal microscopy for intracellular co-localization of proteins. *Methods Mol. Biol.* **261**, 399-410 (2004).

21. Abbe,E. Beitrage zur Theorie des Mikroskops und der mikroskopischen Wahrmehmung. *Arch. F. Mikroskop. Anat* **9**, 413-420 (1873).

22. Hell,S.W., Dyba,M. & Jakobs,S. Concepts for nanoscale resolution in fluorescence microscopy. *Curr. Opin. Neurobiol.* **14**, 599-609 (2004).

23. Hell,S.W. Far-field optical nanoscopy. *Science* **316**, 1153-1158 (2007).

24. Jares-Erijman,E.A. & Jovin,T.M. FRET imaging. *Nat. Biotechnol.* **21**, 1387-1395 (2003).

25. Kenworthy,A.K. Imaging protein-protein interactions using fluorescence resonance energy transfer microscopy. *Methods* **24**, 289-296 (2001).

26. Chien,C.T., Bartel,P.L., Sternglanz,R. & Fields,S. The two-hybrid system: a method to identify and clone genes for proteins that interact with a protein of interest. *Proc. Natl. Acad. Sci. U. S. A* **88**, 9578-9582 (1991).

27. Fields,S. & Song,O. A novel genetic system to detect protein-protein interactions. *Nature* **340**, 245-246 (1989).

28. Massoud,T.F., Paulmurugan,R., De,A., Ray,P. & Gambhir,S.S. Reporter gene imaging of protein-protein interactions in living subjects. *Curr. Opin. Biotechnol.* **18**, 31-37 (2007).

29. Parrish,J.R., Gulyas,K.D. & Finley,R.L., Jr. Yeast two-hybrid contributions to interactome mapping. *Curr. Opin. Biotechnol.* **17**, 387-393 (2006).

30. Mitra,R.D., Silva,C.M. & Youvan,D.C. Fluorescence resonance energy transfer between blue-emitting and red-shifted excitation derivatives of the green fluorescent protein. *Gene* **173**, 13-17 (1996).

31. Miyawaki,A. & Tsien,R.Y. Monitoring protein conformations and interactions by fluorescence resonance energy transfer between mutants of green fluorescent protein. *Methods Enzymol.* **327**, 472-500 (2000).

32. Zacharias,D.A., Violin,J.D., Newton,A.C. & Tsien,R.Y. Partitioning of lipid-modified monomeric GFPs into membrane microdomains of live cells. *Science* **296**, 913-916 (2002).

33. Wallrabe,H. & Periasamy,A. Imaging protein molecules using FRET and FLIM microscopy. *Curr. Opin. Biotechnol.* **16**, 19-27 (2005).

34. Truong,K. & Ikura,M. The use of FRET imaging microscopy to detect protein-protein interactions and protein conformational changes in vivo. *Curr. Opin. Struct. Biol.* **11**, 573-578 (2001).

35. Zhang,J., Campbell,R.E., Ting,A.Y. & Tsien,R.Y. Creating new fluorescent probes for cell biology. *Nat. Rev. Mol. Cell Biol.* **3**, 906-918 (2002).

36. Pfleger,K.D. & Eidne,K.A. Illuminating insights into protein-protein interactions using bioluminescence resonance energy transfer (BRET). *Nat. Methods* **3**, 165-174 (2006).

37. Jensen,A.A., Hansen,J.L., Sheikh,S.P. & Brauner-Osborne,H. Probing intermolecular protein-protein interactions in the calcium-sensing receptor homodimer using bioluminescence resonance energy transfer (BRET). *Eur. J. Biochem.* **269**, 5076-5087 (2002).

38. Xu,Y., Piston,D.W. & Johnson,C.H. A bioluminescence resonance energy transfer (BRET) system: application to interacting circadian clock proteins. *Proc. Natl. Acad. Sci. U. S. A* **96**, 151-156 (1999).

39. Villalobos,V., Naik,S. & Piwnica-Worms,D. Current state of imaging protein-protein interactions in vivo with genetically encoded reporters. *Annu. Rev. Biomed. Eng* **9**, 321-349 (2007).

40. Remy,I. & Michnick,S.W. Mapping biochemical networks with protein-fragment complementation assays. *Methods Mol. Biol.* **261**, 411-426 (2004).

41. Pelletier,J.N., Campbell-Valois,F.X. & Michnick,S.W. Oligomerization domain-directed reassembly of active dihydrofolate reductase from rationally designed fragments. *Proc. Natl. Acad. Sci. U. S. A* **95**, 12141-12146 (1998).

42. Remy,I., Campbell-Valois,F.X. & Michnick,S.W. Detection of protein-protein interactions using a simple survival protein-fragment complementation assay based on the enzyme dihydrofolate reductase. *Nat. Protoc.* **2**, 2120-2125 (2007).

43. Remy,I. & Michnick,S.W. Clonal selection and in vivo quantitation of protein interactions with protein-fragment complementation assays. *Proc. Natl. Acad. Sci. U. S. A* **96**, 5394-5399 (1999).

44. Blakely,B.T. *et al.* Epidermal growth factor receptor dimerization monitored in live cells. *Nat. Biotechnol.* **18**, 218-222 (2000).

45. Rossi,F., Charlton,C.A. & Blau,H.M. Monitoring protein-protein interactions in intact eukaryotic cells by beta-galactosidase complementation. *Proc. Natl. Acad. Sci. U. S. A* **94**, 8405-8410 (1997).

46. Ozawa,T., Nogami,S., Sato,M., Ohya,Y. & Umezawa,Y. A fluorescent indicator for detecting protein-protein interactions in vivo based on protein splicing. *Anal. Chem.* **72**, 5151-5157 (2000).

47. Noren,C.J., Wang,J. & Perler,F.B. Dissecting the Chemistry of Protein Splicing and Its Applications. *Angew. Chem. Int. Ed Engl.* **39**, 450-466 (2000).

48. Hu,C.D. & Kerppola,T.K. Simultaneous visualization of multiple protein interactions in living cells using multicolor fluorescence complementation analysis. *Nat. Biotechnol.* **21**, 539-545 (2003).

49. Grinberg,A.V., Hu,C.D. & Kerppola,T.K. Visualization of Myc/Max/Mad family dimers and the competition for dimerization in living cells. *Mol. Cell Biol.* **24**, 4294-4308 (2004).

50. Demidov,V.V. *et al.* Fast complementation of split fluorescent protein triggered by DNA hybridization. *Proc. Natl. Acad. Sci. U. S. A* **103**, 2052-2056 (2006).

51. Zlokarnik,G. *et al.* Quantitation of transcription and clonal selection of single living cells with beta-lactamase as reporter. *Science* **279**, 84-88 (1998).

52. Hasegawa,S., Jackson,W.C., Tsien,R.Y. & Rao,J. Imaging Tetrahymena ribozyme splicing activity in single live mammalian cells. *Proc. Natl. Acad. Sci. U. S. A* **100**, 14892-14896 (2003).

53. Hasegawa,S., Choi,J.W. & Rao,J. Single-cell detection of trans-splicing ribozyme in vivo activity. *J. Am. Chem. Soc.* **126**, 7158-7159 (2004).

54. Wehrman,T., Kleaveland,B., Her,J.H., Balint,R.F. & Blau,H.M. Protein-protein interactions monitored in mammalian cells via complementation of beta -lactamase enzyme fragments. *Proc. Natl. Acad. Sci. U. S. A* **99**, 3469-3474 (2002).

55. Galarneau,A., Primeau,M., Trudeau,L.E. & Michnick,S.W. Beta-lactamase protein fragment complementation assays as in vivo and in vitro sensors of protein protein interactions. *Nat. Biotechnol.* **20**, 619-622 (2002).

56. Remy,I., Ghaddar,G. & Michnick,S.W. Using the beta-lactamase protein-fragment complementation assay to probe dynamic protein-protein interactions. *Nat. Protoc.* **2**, 2302-2306 (2007).

57. Pichler,A., Prior,J.L. & Piwnica-Worms,D. Imaging reversal of multidrug resistance in living mice with bioluminescence: MDR1 P-glycoprotein transports coelenterazine. *Proc. Natl. Acad. Sci. U. S. A* **101**, 1702-1707 (2004).

58. Ozawa,T., Kaihara,A., Sato,M., Tachihara,K. & Umezawa,Y. Split luciferase as an optical probe for detecting protein-protein interactions in mammalian cells based on protein splicing. *Anal. Chem.* **73**, 2516-2521 (2001).

59. Kaihara,A., Kawai,Y., Sato,M., Ozawa,T. & Umezawa,Y. Locating a protein-protein interaction in living cells via split Renilla luciferase complementation. *Anal. Chem.* **75**, 4176-4181 (2003).

60. Paulmurugan,R., Umezawa,Y. & Gambhir,S.S. Noninvasive imaging of protein-protein interactions in living subjects by using reporter protein complementation and reconstitution strategies. *Proc. Natl. Acad. Sci. U. S. A* **99**, 15608-15613 (2002).

61. Luker,K.E. *et al.* Kinetics of regulated protein-protein interactions revealed with firefly luciferase complementation imaging in cells and living animals. *Proc. Natl. Acad. Sci. U. S. A* **101**, 12288-12293 (2004).

62. Stefan,E. *et al.* Quantification of dynamic protein complexes using Renilla luciferase fragment complementation applied to protein kinase A activities in vivo. *Proc. Natl. Acad. Sci. U. S. A* **104**, 16916-16921 (2007).

63. Remy,I. & Michnick,S.W. Visualization of biochemical networks in living cells. *Proc. Natl. Acad. Sci. U. S. A* **98**, 7678-7683 (2001).

64. Tarassov,K. *et al.* An in Vivo Map of the Yeast Protein Interactome. *Science* (2008).

65. Remy,I. & Michnick,S.W. A cDNA library functional screening strategy based on fluorescent protein complementation assays to identify novel components of signaling pathways. *Methods* **32**, 381-388 (2004).

66. Remy,I., Montmarquette,A. & Michnick,S.W. PKB/Akt modulates TGF-beta signalling through a direct interaction with Smad3. *Nat. Cell Biol.* **6**, 358-365 (2004).

67. von Mering,C. *et al.* Comparative assessment of large-scale data sets of protein-protein interactions. *Nature* **417**, 399-403 (2002).

Chapter 2: Development of an RNA-based detection of endogenous protein-protein interactions

The work discussed in this chapter is unpublished. Initial experiments on the original hammerhead ribozyme greatly benefited from the advice of Ed Curtis (Prof. David Bartel laboratory, MIT Biology Department). Dan Chinnapen, a post-doctoral fellow in our laboratory, assisted with the expression and purification of the T7 RNA polymerase used in the *in vitro* transcription reactions.

Introduction

In this chapter we describe the steps taken towards the development of a new methodology for the detection of endogenous protein-protein interactions (PPIs) in live cells. With this methodology, we hoped to bridge the ability of immunofluorescence (IF) and immunoprecipitation (IP) to detect endogenous proteins and the live-cell compatibility of protein complementation assays (PCAs). Starting from a known hammerhead ribozyme and a previously reported RNA aptamer against the p50 member of the NF*k*B family of transcription factors, we describe our molecular engineering of novel aptamer-ribozyme fusions capable of both binding the target protein and catalyzing the cleavage of a reporter substrate *in vitro*. We also analyze the limitations of the methodology, which explain the failure of the engineered system to detect the homodimerization of the p50 protein. Finally, we discuss potential methodology improvements that could lead to the successful application of this system to the detection of other interacting protein pairs.

Proposed scheme to detect endogenous protein-protein interactions in live cells

Our overall proposed strategy is analogous to the enzyme complementation assays (ECAs). We chose ECAs as our starting platform because they already offer many of the desired properties for an ideal PPI detection method: spatial and temporal resolution, high sensitivity, and live-cell compatibility. In existing ECAs, a reporter enzyme such as β-lactamase[1,2], luciferase[3,4], or dihydrofolate reductase (DHFR)[5,6] is split into two halves, each of which is fused to a protein of interest (Figure 2-1A). In the absence of a PPI, the two enzyme fragments are separated and unable to perform catalysis. In contrast, if the two proteins of interest interact, the two enzyme fragments are brought into close enough proximity so that they can re-combine, thereby reconstituting enzyme activity. This enzymatic activity usually translates into the generation of a fluorescent (e.g. β-lactamase assays) or luminescent (e.g. luciferase assays) signal. ECAs have been used in live mammalian cells[1,4,7] and in animals[3] and have shown greater sensitivity than other PCAs such as the split-GFP[8] system owing to their enzymatic signal amplification.

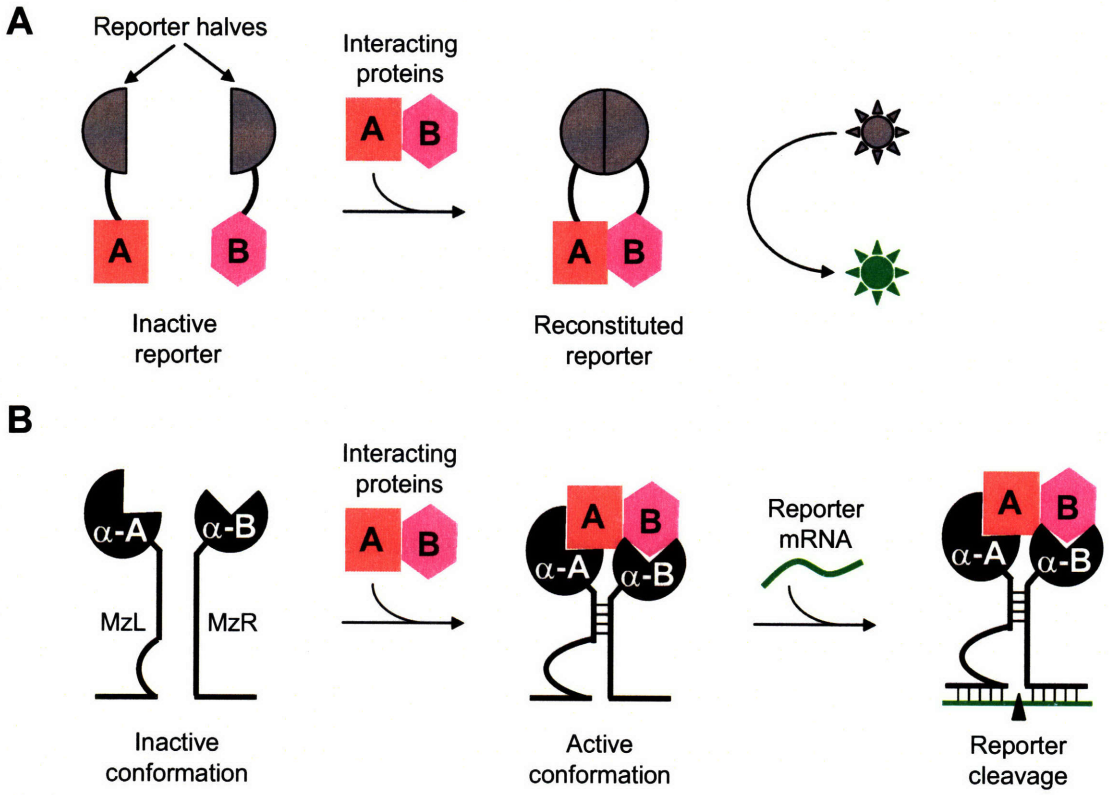


Figure 2-1: Proposed scheme for the detection of endogenous protein-protein interactions in living cells. (**A**) Protein-based detection of recombinant protein-protein interactions via enzyme fragment complementation (ECA). In ECAs, each half of a reporter enzyme (grey semi-circle) is fused to one of the interacting protein partners (A and B). If the proteins A and B interact, the two halves of the enzyme reporter come together and recombine. The reconstituted enzyme is then capable of catalyzing the conversion of a small-molecule substrate (grey star) into a detectable (i.e., fluorescent or luminescent) form (green star). (**B**) RNA-based detection of endogenous protein-protein interactions. In our detection scheme, each half of a dimeric ribozyme (MzL and MzR) is fused to an RNA aptamer against one of the interacting partners (α-A and α-B). Upon interaction of the protein targets, the two halves of the ribozyme come together into an active conformation capable of cleaving its substrate mRNA. Cleavage of the reporter mRNA ultimately results in a decrease in the cellular concentration of the reporter.

We then sought to improve ECAs even further by adapting them to the detection of endogenous PPIs. In order to detect endogenous proteins, the linkage between the reporter fragment and the protein of interest cannot be generated by genetic fusion. One approach could involve fusing each reporter half to an antibody against the target protein. Unfortunately, antibodies are not membrane permeant and are not easily expressed in mammalian cells. In contrast to antibodies, their RNA counterparts, RNA aptamers, can be easily introduced into mammalian cells by transfection. Aptamers are nucleic acid binding species with affinities and specificities that rival those of monoclonal antibodies

(e.g., aptamer dissociation constants are usually in the picomolar to low nanomolar range)[9]. Typically 15–40 nucleotides (nt) long, the nucleotide chain of DNA and RNA aptamers folds in solution into a complex three-dimensional shape, which allows the aptamer to bind tightly against the surface of its target molecule[9]. Since their first inception, RNA aptamers have proven very versatile: using systematic evolution of ligands by exponential enrichment (SELEX) many aptamers have been generated *de novo* against a wide variety of ligands, including small molecules, peptides, and proteins[10]. In general, introduction of RNA sequences into molecular imaging systems is appealing because they offer simpler means for structure and function prediction than proteins.

We thus decided to translate the concept of enzyme complementation into an all-RNA assay. Figure 2-1B depicts the proposed scheme. In this case, the reporter is a heterodimeric ribozyme called the maxizyme (Mz). An RNA sequence selected for high-affinity binding to the target protein, an RNA aptamer, is then used as the scaffold to join the ribozyme to the protein of interest. In this manner, the protein fusions used in ECAs are substituted by an RNA aptamer-ribozyme fusion, an aptazyme, which can both recognize the endogenous protein and catalyze the signal amplification cascade. Because aptazymes are genetically encodable, they can be transiently or stably transfected inside cells and should therefore be able to detect endogenous interactions in a non-invasive manner.

A similar approach to detecting protein-protein interactions using RNA aptamers and hammerhead ribozymes had been previously developed by the Famulok lab (University of Bonn) in collaboration with the Ellington lab (University of Texas)[11]. In their system, interaction of any protein partner with the aptamer-binding protein decoys the protein from the aptamer, which is then capable of hybridizing with an antisense sequence within the ribozyme. In turn, aptamer hybridization induces a conformational change in the ribozyme that renders it catalytically inactive. They showed that this detection scheme was able to report on the binding of the hirudin inhibitor to α-thrombin *in vitro*. Unfortunately, in this system, the aptamer binds to the same protein region as its binding partner, thus acting as a competitive inhibitor. This means that, if the affinity of the aptamer for its target protein is higher than that of the protein-protein interaction, the approach will effectively inhibit the PPI rather than reporting on it. For this reason, and

because of the need for injection of the molecular beacons used as ribozyme substrates, this system would be too invasive and possess too low a sensitivity for low-affinity interactions to be generally applicable *in vivo*.

The following sections describe our choice of a ribozyme and a reporter enzyme for our system.

The maxizyme: a dimeric hammerhead ribozyme

A ribozyme (Rz) is an RNA sequence that acts as a chemical catalyst; naturally occurring ribozymes, for example, catalyze the cleavage or ligation of RNA sequences. Ribozymes are found in most natural organisms and can be classified into cleaving ribozymes, which include self-cleaving RNAs and the trans-cleaving ribonuclease P (RNase P), and splicing ribozymes, which are large RNAs involved in the excision of introns from precursor RNAs[12]. The self-cleaving ribozymes include the ribosomal RNA, the smaller hammerhead, hairpin, hepatitis δ virus (HDV), and *Neurospora* Varkud Satellite (VS) ribozymes[13], and the recently discovered riboswitches such as the bacterial *glmS* riboswitch[14] or the human co-transcriptional cleavage (CoTC) ribozymes[15]. Additionally, many more ribozymes have been engineered in the past 10 years that incorporate novel functionalities, such as C-C bond formation[16], and kinase[17] or polymerase activity[18,19].

For several reasons, we chose the hammerhead ribozyme (HHRz) as the catalytic component of our system (Figure 2-2A). First, HHRz is the smallest naturally occurring ribozyme capable of performing sequence-specific cleavage[13]. Second, owing to its small size, ease of redirection to new substrates, and rapid kinetics, the hammerhead ribozyme is the most widely studied ribozyme and there are many examples of HHRz engineering to allow cleavage of different substrate RNAs *in vitro*. Third, the HHRz has been shown to be, at least partially, active *in vivo*[20,21]. Fourth, although natural HHRzs catalyze self-cleavage, they were, soon after their initial discovery, engineered to perform multiple-turnover trans-cleavage[22], and could thus be capable of signal amplification. Finally, a heterodimeric hammerhead ribozyme had already been developed by the Taira lab (University of Tsukuba, Japan). Although a split ribozyme based on the *Tetrahymena*

group I intron was reported in 2006[23], the former was the only dimeric ribozyme reported at the time when this project was being pursued.

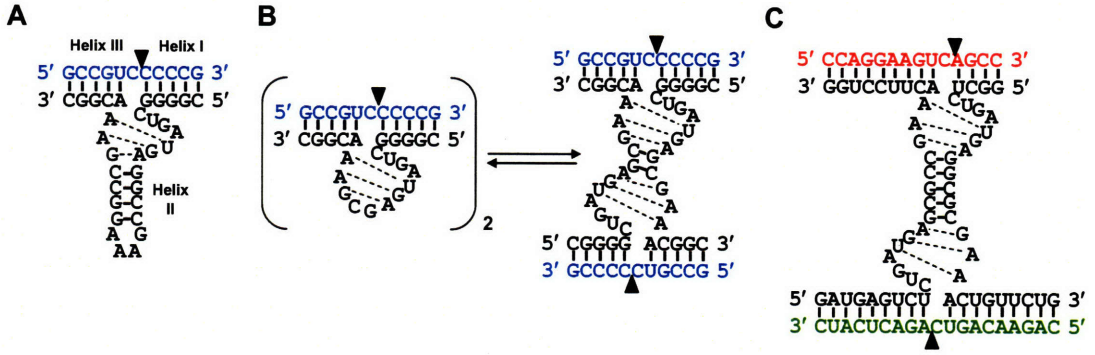


Figure 2-2: Evolution of the hammerhead ribozyme towards the dimeric maxizyme. (**A**) Secondary structure of the minimal hammerhead ribozyme in complex with its substrate (blue). The original HHRz was modified to identify the minimal essential elements for catalytic activity, yielding the minimal HHRz shown here. This minimal trans-cleaving HHRz contains three helical stems – I, II, and III – which flank the nine conserved bases of the catalytic core. Helices I and III are involved in substrate specificity, whereas the catalytic core is involved in the formation of the tertiary structure necessary for cleavage[24]. (**B**) Secondary structure of a minizyme in complex with its substrate (blue). The minizyme results from deleting stem II from the minimal hammerhead ribozyme. The minizyme is inactive as a monomer but has a strong catalytic activity as a dimer. This homodimer is called a maxizyme. (**C**) Secondary structure of a heterodimeric maxizyme resulting from mixing two distinct monomers. The heterodimeric maxizyme has two different substrate-binding arms and is therefore capable of cleaving two different substrates (red and green). Image adapted from Amontov and Taira[25].

The dimeric hammerhead ribozymes developed in the Taira lab, the so-called maxizymes, were discovered when attempting to replace the stem loop II of the original HHRz by a short linker (Figure 2-2B). This shortened HHRz, the minizyme, was found to be inactive as a monomer, but exhibits measurable catalytic activity ($k_{cleav} \sim 2–3\ min^{-1}$) as a dimer[25,26] (Figure 2-2B). By combining two different monomers, one can then design a heterodimer capable of cleaving two different substrates (Figure 2-2C), although the overall catalytic efficiency decreases with respect to the homodimer ($k_{cleav} \leq 0.24\ min^{-1}$)[27]. The heterodimeric maxizymes can also be turned into allosterically regulated ribozymes by deleting one of the catalytic cores and using one of the two substrate-binding regions as sensor arms. In this design, when the sensor arms pair with the correct target sequence, the maxizyme is able to adopt the correct conformation and cleave the target sequence at a different site[28]. Importantly, all these heterodimeric maxizymes have been shown to be catalytically active in cells[29,30] and in animals[31].

Although the Taira lab has not yet developed any fusion of the maxizyme to an aptamer, fusions of the minimal hammerhead ribozyme to RNA aptamers similar to those that are required in our system have been previously engineered. In previously reported aptazymes, the catalytic activity of the HHRz domain is controlled by ligands interacting with the aptamer domain[11,32]. The most useful approach to the design of these allosterically regulated ribozymes is probably the one developed by the Breaker lab (Yale University). In this approach, the RNA aptamer is fused to the stem loop II of a minimal hammerhead ribozyme through a communication module[32]. In the presence of bound analyte, the aptamer induces a conformational change in the catalytic core of the ribozyme, thereby activating (or inhibiting) its catalytic activity. The combined use of rational design and *in vitro* evolution of the communication module provided two allosteric ribozymes with k_{cat} values of ~ 1 min^{-1} and a dynamic range for allosteric activation of 5000-fold in response to cGMP and Co^{2+} (Ref. 32). It remains to be demonstrated whether similar aptamer-ribozyme fusions can be developed based on the heterodimeric maxizyme. It is worth mentioning that, to date, only one naturally occurring regulated ribozyme has been found. In the *glmS* aptazyme, glucosamine-6-phosphate (GlcN6P) serves as a cofactor, which, upon binding to the aptazyme, participates in the cleaving of the mRNA encoding the GlmS enzyme that synthesizes GlcN6P[14].

β-lactamase: a sensitive reporter system

We chose β-lactamase as the reporter system because of its demonstrated high sensitivity (i.e., detection of as little as 100 β-lactamase protein molecules)[33-35], and because its activity can be detected in live cells by fluorescence microscopy. β-lactamases are a family of bacterial enzymes that cleave β-lactams such as penicillins and cephalosporins very efficiently. The TEM-1 β-lactamase from *E. coli* was first conceived as a reporter of gene expression in the laboratory of Roger Tsien (UC San Diego)[35]. Figure 2-3 shows the principles of the β-lactamase reporter system. Tsien and co-workers designed a β-lactamase substrate in which two fluorophores (a coumarin donor and a fluorescein acceptor) are attached to the 7 and 3′ positions of a cephalosporin

heterocycle. In this configuration, the two fluorophores are close enough to exhibit efficient fluorescence resonance energy transfer (FRET). However, β-lactamase cleavage of the cephalosporin results in elimination of the fluorophore from the 3′ position, disrupts FRET, and reestablishes fluorescence emission by the coumarin donor, thus resulting in an increase in the cellular blue-to-green emission signal ratio. Alternatively, fluorogenic substrates have also been designed, which transform from non-fluorescent to brightly fluorescent (~ 153-fold enhancement in fluorescence intensity) after β-lactamase hydrolysis[36]. Since its inception in 1998, the β-lactamase reporter system has been extensively used in protein complementation assays to detect protein-protein interaction in cells[1,2,7] and, owing to its high sensitivity, has enabled single-cell monitoring of the self-[33] and trans-splicing[34] activities of the *Tetrahymena* group I intron ribozyme.

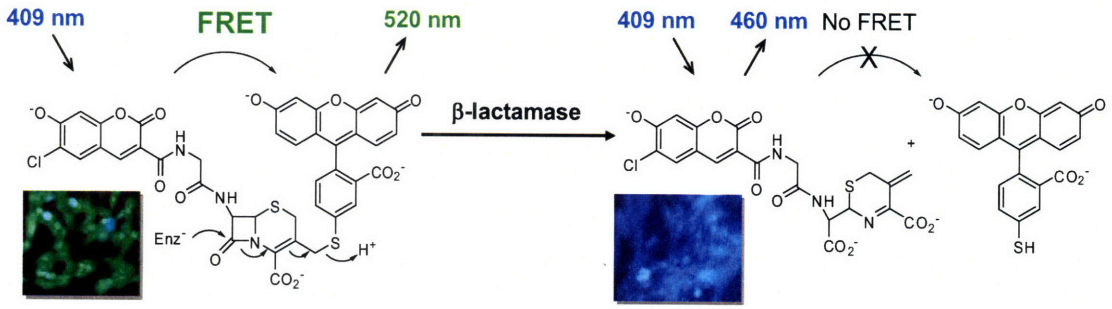


Figure 2-3: The β-lactamase reporter system. In the intact substrate, excitation of a coumarin donor (left) at 409 nm leads to fluorescence resonance energy transfer (FRET) to the fluorescein acceptor (right) and re-emission of green light. β-lactamase (Enz) catalyzes the hydrolysis of a cephalosporin β-lactam ring to generate two products in which the donor and acceptor are separated. The donor then emits blue fluorescence, whereas the acceptor is quenched. Scheme adapted from Zlokarnik *et al*[35].

In retrospect, the β-lactamase system does not seem the best option because it has been shown that its fluorogenic substrate is not taken up by many cell types[2]. In contrast, the use of luciferase as a reporter enzyme retains the high sensitivity of β-lactamase but it has the added advantage that it has been robustly demonstrated both in living cells[3,4] and in whole animals[37]. Furthermore, the use of a bioluminescent signal instead of fluorescence eliminates the background introduced by the auto-fluorescence of biological samples.

Proof-of-principle: homodimerization of the p50 member of the NFκB family of transcription factors

Our starting point in the search for an initial proof-of-principle system was the public aptamer database maintained by the Ellington lab (Austin, Texas)[38]. Among all reported RNA aptamers selected against protein targets, we chose the aptamer against the p50 member of the nuclear factor κB (NFκB) family of transcription factors[39] because it had been extensively characterized both *in vitro* and *in vivo*. The 29-nt aptamer had been initially selected *in vitro*[39] and was later optimized for *in vivo* binding using yeast genetic selections[40]. This was an important consideration for our system, since most other reported aptamers are solely selected *in vitro*, and are therefore not guaranteed to fold properly and to retain binding specificity for their targets when expressed in living cells. Despite their lack of sequence identity, the RNA aptamer displays similarly high binding affinity (K_d = 5.4 ± 2.0 nM) as the native consensus DNA target for $(p50)_2$ and it is also highly selective for the p50 member of the family[40] (i.e., it does not recognize p65). Furthermore, the crystal structure of the aptamer-protein complex had been solved and showed independent binding of one RNA aptamer molecule to each monomer of the p50 homodimer (Figure 2-4)[41]. This is essential to our system, which requires aptamer binding to be independent of the protein-protein interaction. Importantly, this aptamer was shown to retain its binding ability in yeast cells, where it was able to efficiently block DNA binding to $(p50)_2$[40,42].

NFκB is a family of transcription factors with a conserved DNA-binding domain and a dimerization region called the Rel homology region (RHR)[43]. The five known members of the family (p50, p52, p65, c-Rel, and RelB) dimerize to form several homo- and heterodimers, which modulate transcription by binding to the regulatory DNA domains (κB-DNA sites) of numerous genes involved in stress responses such as the immune response, inflammatory response, and programmed cell death[44]. The diversity of the different homo- and heterodimers, each one with distinct binding affinities for the different target DNAs and for the members of the inhibitor protein family IκB, forms the basis of the tightly regulated NFκB signaling[45].

Thus, the existence of a selective and high-affinity aptamer for p50 combined with the wealth of information available on the homodimerization of p50 to form an active transcriptional regulatory complex, led us to choose this as our model system for our proof-of-principle aptazyme studies.

Figure 2-4: Crystal structure of the NF-κB $(p50)_2$ complexed to a high-affinity RNA aptamer. The 2.45Å resolution x-ray crystal structure reveals that two RNA molecules (white backbone with color-coded nucleotides) bind independently of each other to one p50 molecule (red and blue) within the $(p50)_2$ homodimer. The crystal structure shows that the aptamer binds through the N-terminal domain of the protein (bottom), and the short linker connecting the N-terminal domain and the dimerization domain (top). Each p50 monomer uses the same surface to recognize the distorted RNA major groove as observed in the κB DNA/p50 complex[46]. For this crystal structure, only the Rel homology region of p50 was used. This image was generated from the original PDB 1OOA file using DS ViewerPro software.

Results

The hammerhead ribozyme can be modulated to cleave the β-lactamase mRNA

Our initial work focused on independently testing the two major components of the system: the ribozyme and the aptamer. We started by modulating the minimal hammerhead ribozyme (HHRz) to cleave the β-lactamase mRNA. Our initial strategy for generating ribozyme transcripts was based on PCR amplification of a template plasmid containing the HHRz downstream of a $tRNA^{Val}$ promoter, used to enhance ribozyme stability and cytosolic export in cells[47], and upstream of a helicase-recruiting sequence, used to unwind the substrate mRNA thus improving ribozyme cleavage activity[48]. The original plasmid containing a HHRz targeted against the luciferase mRNA (pUC-tRNA-LucRz-CTE) was a gift from Prof. Taira (University of Tsukuba, Japan). We designed

HHRz targeting three different sequences within the β-lactamase gene, by changing the nucleotide sequence of the substrate binding arms (helixes I and III in Figure 2-2A). We chose β-lactamase sequences 1–3 (Figure 2-5), which encompass a GUC (1 & 2) or a CUC (3) triplet. Although HHRzs can cleave any RNA substrate that contains an NUH triplet, where N is any nucleotide and H is A, C, or U, it has been shown that the efficiency of cleavage is higher for GUC and, to a lesser extent, for CUC[49]. Additionally, it is recommended to target sequences that lie either within the first 500 bases of the transcript or at the very end, as these are less likely to be buried within mRNA tertiary structures[50].

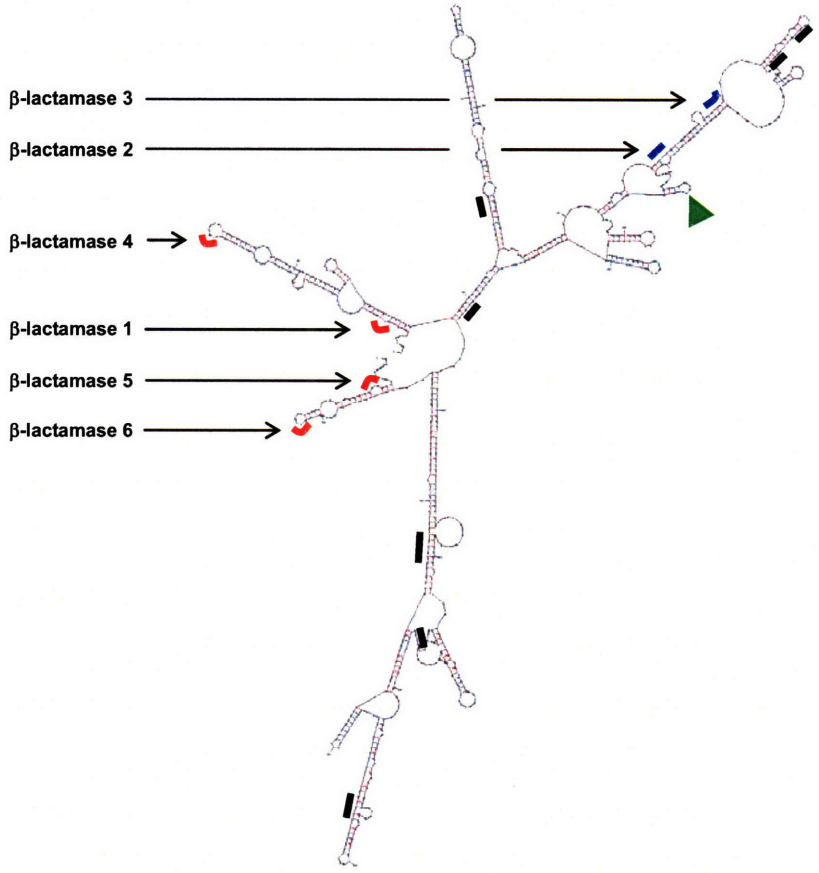


Figure 2-5: Secondary structure of the β-lactamase mRNA as predicted by Mfold. Marked with lines are all the GUC or CUC triplets present in the transcript. Over the course of this project, several ribozymes were designed to target triples 1–6. Triplets marked in black lie within stable stem structures and are likely to be inaccessible, so they were not tested. Triplets in blue (β-lactamase 2 and 3) were tested but no activity was observed when the hammerhead was incubated with a transcript containing the first 380 bases of the mRNA (marked by the green arrowhead). HHRzs designed to cleave the triplets marked in red (β-lactamase 1 and 4–6) showed varying degrees of cleavage activity.

HHRzs targeting β-lactamase sequences 1–3 were first cloned into pUC-tRNA^{Val}-CTE, then PCR-amplified using a forward primer containing a T7 promoter sequence, and finally transcribed using T7 RNA polymerase. The final size of the transcripts was ~ 200 nt for the ribozymes and ~ 400 nt for the β-lactamase substrate. We did not observe any cleavage activity when the HHRzs were tested in an *in vitro* reaction with ^{32}P end-labeled β-lactamase RNA (data not shown). Changing the reaction conditions, such as pH, time, and ribozyme or substrate concentration did not result in any detectable cleavage activity.

We then changed our protocol for generating transcripts. Following the advice of Dr. Ed Curtis (at the time, a graduate student in the Bartel laboratory at MIT), we ordered short DNA oligonucleotides (~ 40–80 nt) containing the HHRz or substrate sequence followed by a 20-nt T7 promoter sequence needed for recognition by T7 RNA polymerase. Transcription yields were a lot higher for the short oligos than they were for the long PCR products. Furthermore, the short HHRz transcript targeting the β-lactamase target sequence 1 was able to efficiently cleave the 400 nt β-lactamase transcript (Figure 2-6). We did not see any activity for HHRzs 2 and 3, perhaps because their target sequences lie within more stable stems or are buried within the RNA tertiary structure.

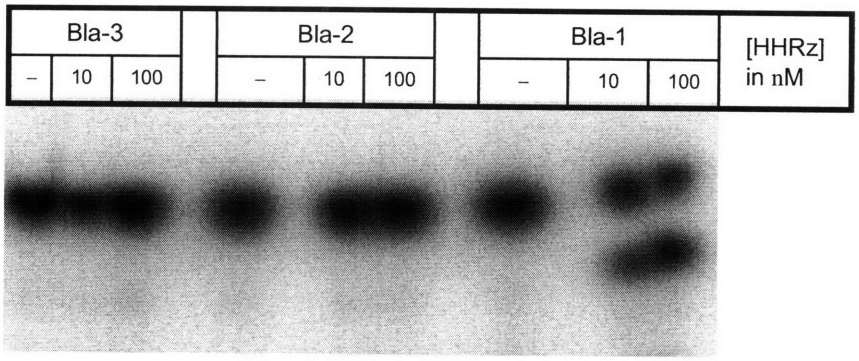


Figure 2-6: The hammerhead ribozyme can be modulated to cleave the β-lactamase mRNA. *In vitro* activity assay of three hammerhead ribozymes targeting β-lactamase target sequences 1–3 (see Figure 2-5). Either 10 or 100 nM HHRz was incubated with ~ 1 nM end-labeled 400 nt β-lactamase transcript for 5 minutes at 37 °C. The reaction products were run on a 10% denaturing acrylamide gel. Negative control reactions (– lanes) were performed in the absence of RNA enzyme.

The heterodimeric maxizyme is active *in vitro* at concentrations higher than 10 nM

In parallel, we started testing the *in vitro* activity of the heterodimeric maxizyme (Mz). Among the several published maxizymes, we chose the heterodimeric Mz that targets the HIV-1 Tat mRNA (Figure 2-2C) because it had been shown to retain its catalytic activity even in the presence of just one substrate[27]. That is, cleavage of the S-19 substrate (red in Figure 2-2C) did not depend on the presence of the S-14 substrate (green in Figure 2-2C) so long as high Mg^{2+} concentrations were included in the *in vitro* reaction. Presumably, high concentrations of Mg^{2+} ions were able to stabilize the dimeric ribozymes in the absence of the second substrate[26] *in vitro*, whereas other factors are responsible for stabilization *in vivo*, where the Mg^{2+} concentration is significantly lower[51]. This was an important consideration in our case because fusion of the RNA aptamers to the maxizyme may have sterically hindered binding of the second substrate.

We therefore prepared transcripts of the left monomer (MzL), the right monomer (MzR), and the S-19 substrate starting from the corresponding short DNA oligos as previously described (see Experimental section). As expected, we only observed ribozyme cleavage activity in the presence of both monomers. Figure 2-7A shows that the S-19 substrate remains intact after 5 minutes incubation with only MzL or only MzR, but ~ 10% cleavage is observed when both monomers are present.

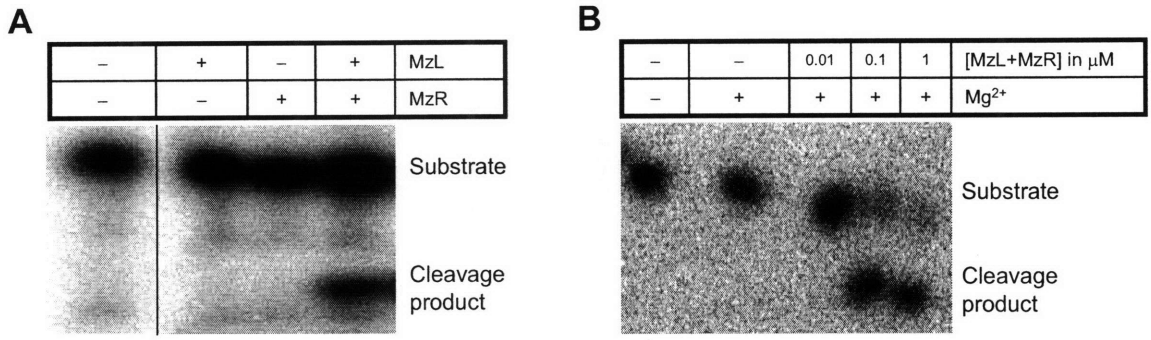


Figure 2-7: *In vitro* cleavage activity of the heterodimeric maxizyme. (**A**) Maxizyme activity requires presence of both maxizyme monomers. 1 nM end-labeled S-19 substrate was incubated for 5 minutes in the absence of RNA enzyme (lane 1) or in the presence of 100 nM of each or both Mz monomers (lanes 2, 3, and 4), and the reaction products were run on a 20% denaturing acrylamide gel. The intact substrate is 19 nt long, and the 5′ cleavage product is 10 nt. (**B**) The maxizyme is active at concentrations higher than 10 nM. The *in vitro* activity assay was performed as in A except that the reaction was incubated for 1 hour with varying concentrations of the maxizyme monomers.

We next sought to find an estimate of the minimum concentration at which the maxizyme is active *in vitro*. This is an important parameter of our PPI detection strategy, which requires that there is no ribozyme activity in the absence of a protein-protein interaction. We performed the *in vitro* activity assay at various concentrations of Mz and observed no detectable cleavage activity when the substrate was incubated in the presence of 10 nM Mz even after one hour reaction (Figure 2-7B). In contrast, incubation with maxizyme concentrations higher than 100 nM resulted in ~ 90% cleavage. It is worth noting that, because assays were performed under single turnover conditions, the extent of cleavage reflects the percentage of maxizyme molecules that are adopting a catalytically active conformation. The results shown in Figure 2-7B are in accordance with the published kinetic parameters for this maxizyme[27]. With a K_M for dimerization of 220 nM (k_{obs} = 0.24 min^{-1}), it is expected that, at 10 nM maxizyme, only ~ 0.4 nM will be in dimeric form, and, of those, only ~ 10% will be in a catalytically active configuration at any given moment. In contrast, at 100 nM maxizyme, ~ 25 nM will be in dimeric form, and even if only 10% is catalytically active, the concentration of active ribozyme is higher than the total concentration of substrate.

Generating an aptazyme that displays binding to p50 and ribozyme catalytic activity

Initial aptazyme fusions did not bind p50

With an active maxizyme in hand, we next turned our attention to the aptamer component of our system. The first step on testing the aptamer was the expression of the wild type (WT) p50 protein and a non-dimerizing mutant (Y267D/L269D) to serve as a negative control for protein-protein interaction. Plasmids for both proteins were obtained from Dr. Gourisankar Ghosh and encoded truncated versions of the proteins with just the Rel homology region (RHR) domains (residues 39–364). The non-dimerizing mutant was N-terminally His-tagged, and was purified by affinity chromatography. The WT protein was un-tagged, and was therefore purified by two subsequent anion and cation exchange chromatographic steps following a published protocol[39]. Both proteins expressed well in *E. coli* BL21 and we were able to routinely obtain purified protein at concentrations ~ 60 μM.

We prepared the aptamer by transcription of the corresponding DNA oligo and we end-labeled it using [γ-^{32}P] ATP as described in the Experimental section. Aptamer binding to p50 was tested by a nitrocellulose filter binding assay. We observed a fractional saturation of aptamer binding of 0.83 ± 0.05 at 50 nM p50 (Figure 2-8B, first bar, and Table 2-1), which is in accordance with the previously reported aptamer affinity[40]. To further confirm the validity of our assay conditions, we prepared point mutations of the aptamer at positions that had been previously postulated to be essential for protein binding[41]. Figure 2-8B shows that all three mutations impaired aptamer binding activity, with mutant 1 (A11:U20 to C11:G20) resulting in lowest binding. Figure 2-8A shows the aptamer secondary structure and highlights its main contacts with p50.

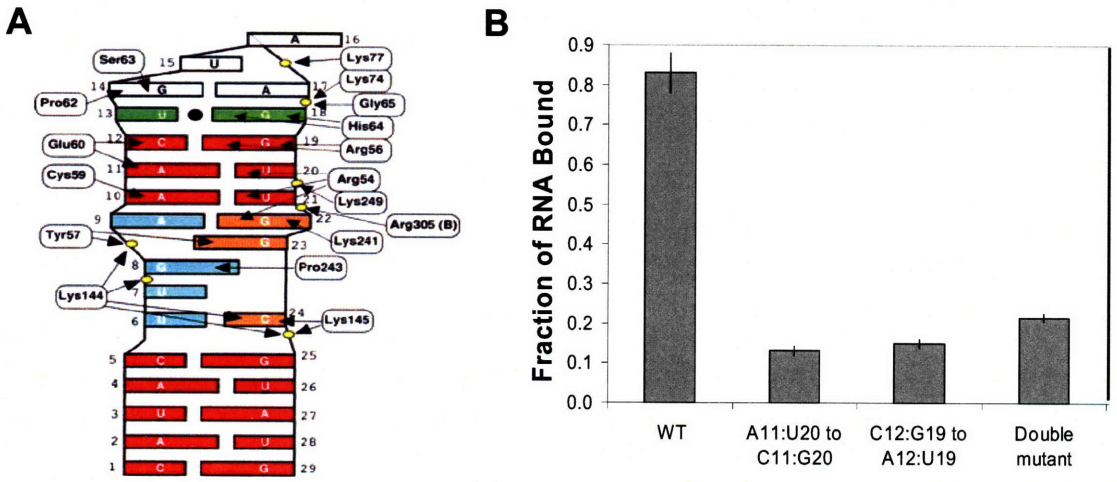


Figure 2-8: Aptamer binding to p50. (**A**) Summary of contacts between the RNA aptamer and p50 as seen in the crystal structure. Watson-Crick base pairs are indicated in red. Residues predicted to form an internal loop are colored cyan (5′) and orange (3′). Yellow ovals represent phosphates in the RNA backbone contacted by p50. Figure taken from Huang et al[41]. (**B**) Fractional saturation of aptamer binding to p50. Aptamer binding to p50 was tested using a nitrocellulose filter binding assay with 0.9 nM end-labeled RNA aptamer and 50 nM p50. Each data point represents the average of three independent experiments. Error bars, 1 s.d. The mutations were introduced in pairs in an attempt to impair protein binding without disrupting aptamer secondary structure.

The first generation of aptazyme fusions was constructed by directly joining the aptamer to the 3′ end of the left monomer of the maxizyme or the 5′ end of the right monomer without any extra linkers. Unfortunately, although catalytically active (data not shown), both fusions displayed a marked decrease in binding to p50. Table 2-1 shows that the fractions of left and right monomers of the aptazyme fusion (called Az-0) bound

to p50 were 0.15 ± 0.01 and 0.17 ± 0.02, respectively. These numbers are comparable to those obtained for the non-binding aptamer mutants (e.g., the A12:U19 mutant aptamer displays a 0.15 ± 0.01 fractional saturation at 50 nM p50). This loss of binding capability can be reasoned based on the predicted secondary structures of the fusions. Figure 2-9 shows that the aptamer secondary structure (Figure 2-9A) is disrupted in the aptazyme fusions because of base-pairing between the aptamer and the maxizyme domains (Figure 2-9B and C).

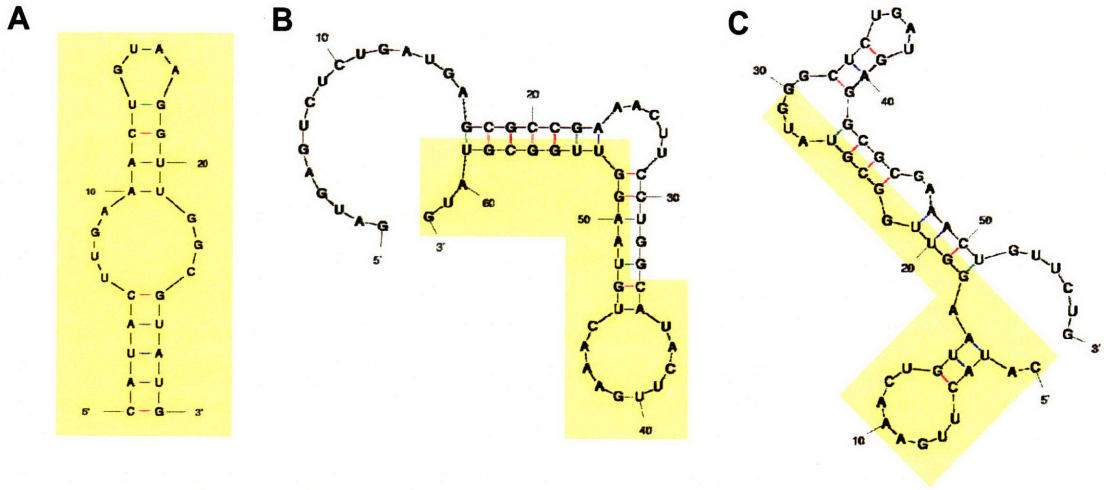


Figure 2-9: Predicted secondary structure of the α-p50 aptamer and the first generation of aptazyme fusions. (**A**) Aptamer against p50. (**B**) Left monomer of the aptazyme fusion (AzL-0). (**C**) Right monomer of the aptazyme fusion (AzR-0).

Extra complementary strands in the aptazyme stabilizes the aptamer and ribozyme domains

We then sought to improve the stability of the aptamer fold in the aptazyme fusions. We pursued two different strategies. First, we introduced two or four extra base pairs in the aptamer stem to additionally stabilize its fold (Figure 2-10, AzR-1 and 2, and AzL-3 and 4). Second, we added extra nucleotides at the 3′ end of the aptamer in AzL, complementary to the secondary substrate binding arm of MzL. These extra nucleotides (Figure 2-10, AzL-1) would play a double role: they prevent formation of incorrect secondary structure due to mispairing of maxizyme and aptamer bases, and they mimic the presence of a second maxizyme substrate, thereby stabilizing the maxizyme active

conformation. Finally, a combination of both strategies was also tested (the combination of AzL-1 and 3 was called AzL-2).

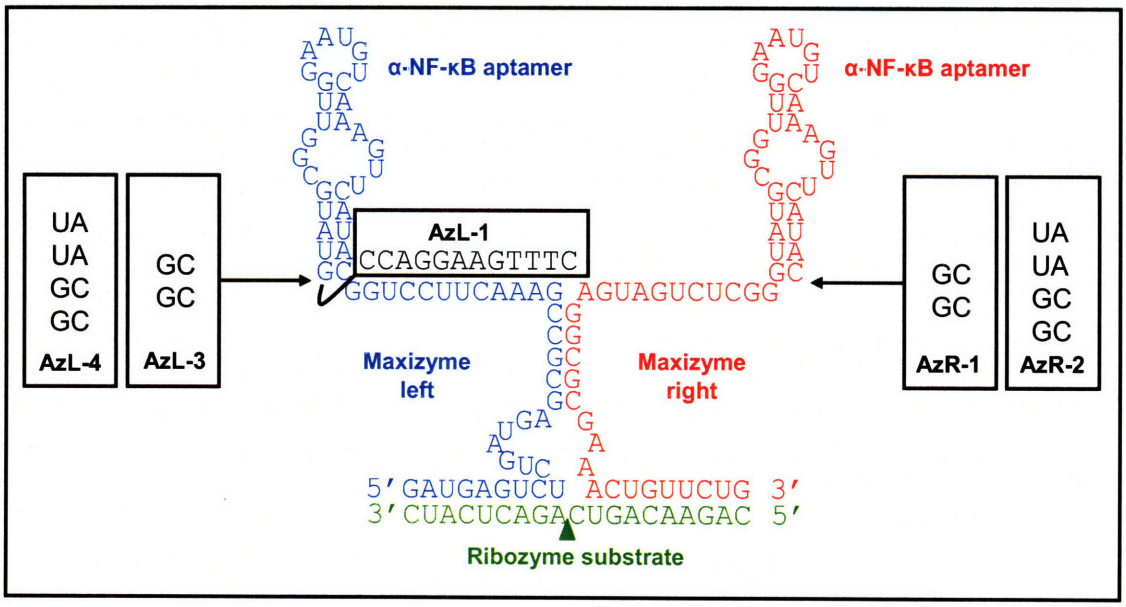


Figure 2-10: Nucleotide sequence of the aptazyme fusions. The RNA aptamer against p50 was fused to the 3′ end of the left monomer of the dimeric maxizyme (MzL) to afford the left monomer of the aptazyme (AzL, blue) or to the 5′ end of the right monomer of the maxizyme (MzR) to create AzR (red). The ribozyme substrate is represented in green. The nucleotides boxed and in black were later added in order to promote independent folding of the aptamer and ribozyme domains of the aptazyme fusions.

We transcribed the DNA oligos, end-labeled the transcripts with [^{32}P-γ] ATP, and tested the different constructs for binding to p50. Table 2-1 summarizes the results. Binding was recovered for all the constructs tested, although to different extents. For example, addition of two or four extra base pairs at the end of the aptamer stem (AzL-3 and AzL-4) resulted in significantly higher fractional saturations (0.63 ± 0.01 and 0.58 ± 0.02, respectively) than that of AzL-0 (0.15 ± 0.02); but still lower than the fractional saturation of the original aptamer (0.83 ± 0.05). In contrast, the combination of aptamer stem stabilization and extra base-pairing of the MzL substrate binding arm (AzL-2) resulted in a fractional saturation of 0.85 ± 0.03, similar to that obtained for the aptamer alone. As expected, the differences obtained by adding two or four extra base pairs to the aptamer stem in AzR were not dramatic, although better binding was observed for AzR-2, which incorporated four extra base pairs.

Table 2-1: Summary of binding capabilities of the aptamer, maxizyme, and aptazyme fusions towards the p50 protein. Binding tests were all performed using the nitrocellulose binding assay with 0.9 nM RNA and 50 nM protein. The results are given as the fraction of total RNA bound to p50. Unless otherwise noted, the maxizyme domain of the aptazyme fusions contained five base pairs in the dimerization domain, like the one shown in Figure 2-2C. The fractional saturation for a maxizyme and an aptazyme with only two bases pairs in the dimerization stem is also shown (marked with an asterisk *). The complete sequences of all constructs can be found in the Experimental section (Table 2-3). Each data point represents the average of three independent experiments. Error bars, 1 s.d. The numbering of the aptazyme fusions follows the convention described in Figure 2-10.

Construct				**Binding to p50 (as fractional saturation)**	
Name		*Ribozyme substrate*	*WT or mutant aptamer*	*Left*	*Right*
Aptamer		NA	WT	0.83 ± 0.05	
			C11:G20 mutant	0.13 ± 0.01	
			A12:U19 mutant	0.15 ± 0.01	
			Double mutant	0.22 ± 0.03	
Maxizyme		HIV-S19	NA	0.28 ± 0.08	0.42 ± 0.05
		β-lactamase	NA	0.21 ± 0.03	0.18 ± 0.02
		β-lactamase*	NA	0.19 ± 0.01	0.14 ± 0.02
Aptazyme	**Az-0**	HIV-S19	WT	0.15 ± 0.02	0.17 ± 0.02
	Az-1		WT	0.79 ± 0.02	0.61 ± 0.06
	Az-2		WT	0.85 ± 0.03	0.72 ± 0.02
			C11:G20 mutant	0.55 ± 0.02	0.49 ± 0.02
			A12:U19 mutant	0.50 ± 0.03	0.45 ± 0.09
			Double mutant	0.53 ± 0.05	0.44 ± 0.01
	Az-3		WT	0.63 ± 0.01	-
	Az-4		WT	0.58 ± 0.02	-
	Az-2.1	β-lactamase	WT	0.69 ± 0.01	0.75 ± 0.02
			C11:G20 mutant	0.19 ± 0.02	0.12 ± 0.01
	Az-2.1*		WT	0.38 ± 0.03	0.71 ± 0.03
			C11:G20 mutant	0.14 ± 0.02	0.13 ± 0.02

We next prepared the point mutations in the aptamer domain of the two best aptazyme constructs AzL-2 and AzR-2. Surprisingly, we found that none of the three mutations tested eliminated binding. The fractional saturation of the aptazyme was reduced to ~ 0.50 for both AzL and AzR, which was significantly higher than the residual binding seen for the mutant aptamers (ranging from 0.13–0.22). This led us to hypothesize that perhaps the maxizyme domain had, by itself, certain affinity for p50. To test this hypothesis, we estimated the fraction of maxizyme alone that bound to p50 using the same nitrocellulose binding assay. We observed that, indeed, the maxizyme alone displayed significant binding to p50 with fractional saturations of 0.28 ± 0.08 for MzR (5 bp, S-19 substrate) and 0.42 ± 0.05 for MzL (5 bp, S-19 substrate). We also tested the specificity of the binding, and found that none of the monomers (MzL or MzR) bound to the unrelated bovine serum albumin (BSA) or to cyan fluorescence protein (CFP). It is worth noting that binding of the mutant aptazyme fusions was higher than that of the mutant aptamer or the maxizyme by themselves. We hypothesized that this was caused by binding cooperativity between the two domains within the context of the fusions, although initial tests did not seem to confirm this hypothesis (data not shown).

Changing substrate specificity minimizes non-specific binding to p50

In order to find out the reason why the maxizyme monomers were binding to p50 in such a specific manner, we examined their nucleotide sequences and predicted secondary structures. Table 2-2 compares the nucleotide sequence of the α-p50 RNA aptamer with that of the maxizyme targeting the S-19 substrate. It was striking to find that the maxizyme sequences shared as many as 8 nt (MzL-S19) or 9 nt (MzR-S19) with the aptamer. The predicted secondary structures showed that, although not the lowest energy structures, both the MzL-S19 and MzR-S19 could fold into alternate structures with fairly low folding energy, and which closely resembled the aptamer fold (data not shown). We then sought to eliminate the common sequences by changing the ribozyme substrate.

Previously, we had found that the hammerhead ribozyme could be modulated to cleave the β-lactamase mRNA at specific sites (see first section of Results). Based on those previous experiments, we examined the activity of four different maxizymes

modulated to cleave at the β-lactamase 1 site or at three other sites (β-lactamase 4–6) that were expected to be accessible for cleavage (see Figure 2-5).

Table 2-2: Nucleotide sequence comparison for the α-p50 RNA aptamer and several maxizymes differing in their target sequences. MzL and MzR refer to the left and right monomers of the heterodimeric maxizyme, respectively. S-19 refers to the 19-mer substrate targeted by the original maxizyme developed in the Taira lab[27]. β-lact 1 and 4–6 refer to different cleavage sites within the β-lactamase mRNA; the numbering is in accordance with that of Figure 2-5. Nucleotides underlined represent the substrate binding arms of the ribozyme, whereas nucleotides in bold highlight the catalytic core. The nucleotides in red are the ones previously postulated to be involved in aptamer binding to p50[41]. Among those, bold nucleotides mark the ones that we have mutated to create the non-binding aptamer mutants (A11:U20 and C12:G19). Highlighted in grey is the aptamer nucleotide sequence that was also found in the maxizyme constructs. Nucleotides in blue represent extra mutations introduced to further differentiate the aptamer and nucleotide sequences.

Construct		RNA Sequence
Aptamer		CAUACUUGAAACUGUAAGGUUGGCGUAUG
MzL	S-19	GAUGAGUCU**CUGAUGA**GCGCCGAAACUUCCUGG
	β-lact. 1	AAUAAGGGC**CUGAUGA**GCGCCGAAGUCUCCUGG
	β-lact. 4	GCGUUUCUGGGU**CUGAUGA**GCGCCGAAGUCUCCUGG
	β-lact. 5	GGGAGGGCU**CUGAUGA**GCGCCGAAGUCUCCUGG
	β-lact. 6	CCATAGUUGCCU**CUGAUGA**GCGCCGAAGUCUCCUGG
MzR	S-19	GGCUCUGAUGAGGCGC**GAA**ACUGUUCUG
	β-lact. 1	**AAG**UCU**ACA**GAGGCGC**GAA**ACACGGAAA
	β-lact. 4	**AAG**UCU**ACA**GAGGCGC**GAA**AGCAAAAACAGG
	β-lact. 5	**AAG**UCU**ACA**GAGGCGC**GAA**ACCAUCUGG
	β-lact. 6	**AAG**UCU**ACA**GAGGCGC**GAA**ACUCCCCGUCG

All four cleavage sites introduced certain degree of dissimilarity between the aptamer and MzR, but had no effect on MzL (see underlined nucleotides in Table 2-2). In order to change MzL, we then introduced an ACU to GCU mutation (blue nucleotides in Table 2-2). We also introduced additional mutations in MzR (blue nucleotides in Table 2-2) to avoid some potential non-productive interactions with the left monomer that had been predicted by the folding algorithm. We must clarify that all these mutations were

introduced at the maxizyme secondary binding arms, which do not have any catalytic or substrate binding function, and should therefore not affect the catalytic activity of the ribozyme. We tested the four constructs for binding to p50 and we found that, although reduced, the changes in the binding properties were not as pronounced as we had anticipated. We observed fractional saturations ranging from 0.1–0.4 (data not shown), which resemble those obtained for the original maxizyme (0.28 for MzL and 0.42 for MzR). Only the constructs targeting β-lactamase 6 displayed significantly reduced background on both monomers with a fractional saturation of 0.21 ± 0.03 for MzL and of 0.18 ± 0.02 for MzR (Table 2-1, Maxizyme with β-lactamase substrate). Although we could not rationalize why this construct but not the others resulted in decreased binding, we decided to move forward and incorporate the new maxizyme into an aptazyme fusion.

We were very happy to find that the new aptazyme fusion (Az-2.1) displayed improved binding behavior. We estimated that the left and right monomers bound the protein with a fractional saturation of 0.69 ± 0.01 and 0.75 ± 0.02, respectively. Furthermore, point mutations in the aptamer domain were able to decrease aptazyme binding to levels similar to the aptamer alone (Table 2-1). In retrospect, the optimization process may have been more effective had we estimated the signal-to-background ratio (i.e., binding of the WT *versus* the point mutation) instead of just binding of the WT construct from the very beginning. For example, AzL-1, which displayed a fractional binding saturation of 0.79 ± 0.02, was initially discarded because of low binding, whereas the final AzL-2.1 optimized monomer showed a 0.69 ± 0.01 fractional saturation.

We next confirmed that the optimized aptazyme fusions retained cleavage activity, and estimated the maximum concentration at which they showed no detectable activity. The results are shown in Figure 2-11. We found that, similar to the original maxizyme (Figure 2-7B), the aptazyme fusion retained ribozyme activity, and that detectable cleavage was observed at concentrations higher than 10 nM in the absence of a protein-protein interaction.

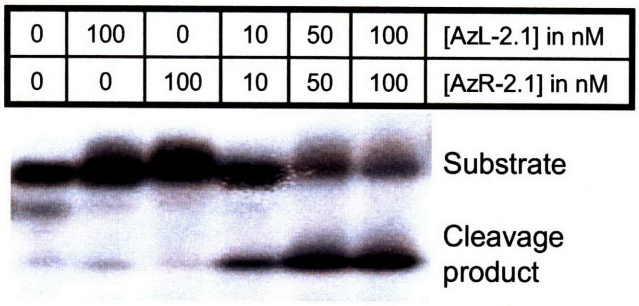


Figure 2-11: *In vitro* cleavage activity of the optimized aptazyme fusion against a target sequence in the β-lactamase mRNA. A 24-mer long β-lactamase-6 substrate was combined with increasing amounts of RNA enzyme, incubated at 37 °C for 1 hour, and the reaction products run in a 20% denaturing acrylamide gel. Negative controls are also included with no enzyme added (lane 1), or with just the left (lane 2) or right monomers (lane 2).

To further characterize the optimized aptazyme, we estimated the equilibrium dissociation constants for p50 binding. We plotted the binding isotherms for AzL and AzR binding to increasing amounts of protein, and obtained a dissociation constant of 32.5 ± 5.1 nM for AzL and 21.5 ± 2.4 nM for AzR, both slightly higher than the reported K_d for the aptamer (5.4 ± 2 nM)[40]. These binding affinities were in accordance with the values of fractional saturation previously obtained (Table 2–1).

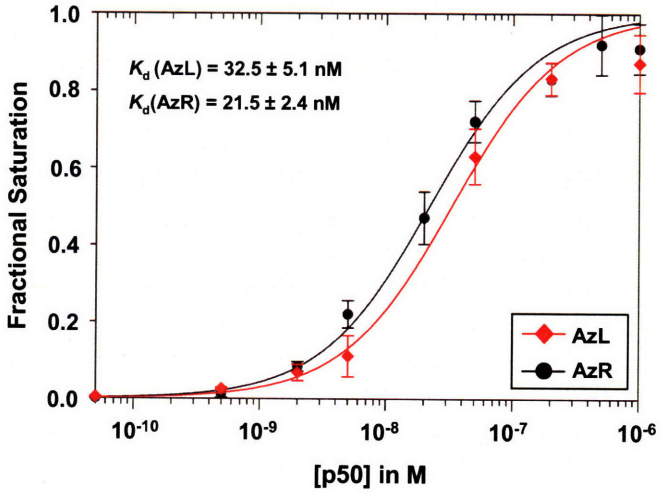


Figure 2-12: Estimation of equilibrium dissociation constants for p50 binding to the RNA optimized aptazyme fusions. Nitrocellulose filter binding assays were performed with radiolabeled left (AzL) or right (AzR) aptazyme monomer in the presence of increasing concentrations of p50 protein. The fractional saturation of the radiolabeled probe is plotted as a function of protein concentration for AzL (♦) or AzR (•). Error bars indicate standard deviation. The indicated equilibrium dissociation constant estimates were determined by curve fitting to the equation described in the Experimental section.

Finally, Figure 2-13 compares the predicted secondary structures of the original and the improved aptazyme fusions. The predicted structures, together with the positive results for binding and cleavage activities, confirmed that the optimization process finally yielded a fusion construct, where both domains can fold and function independently of each other.

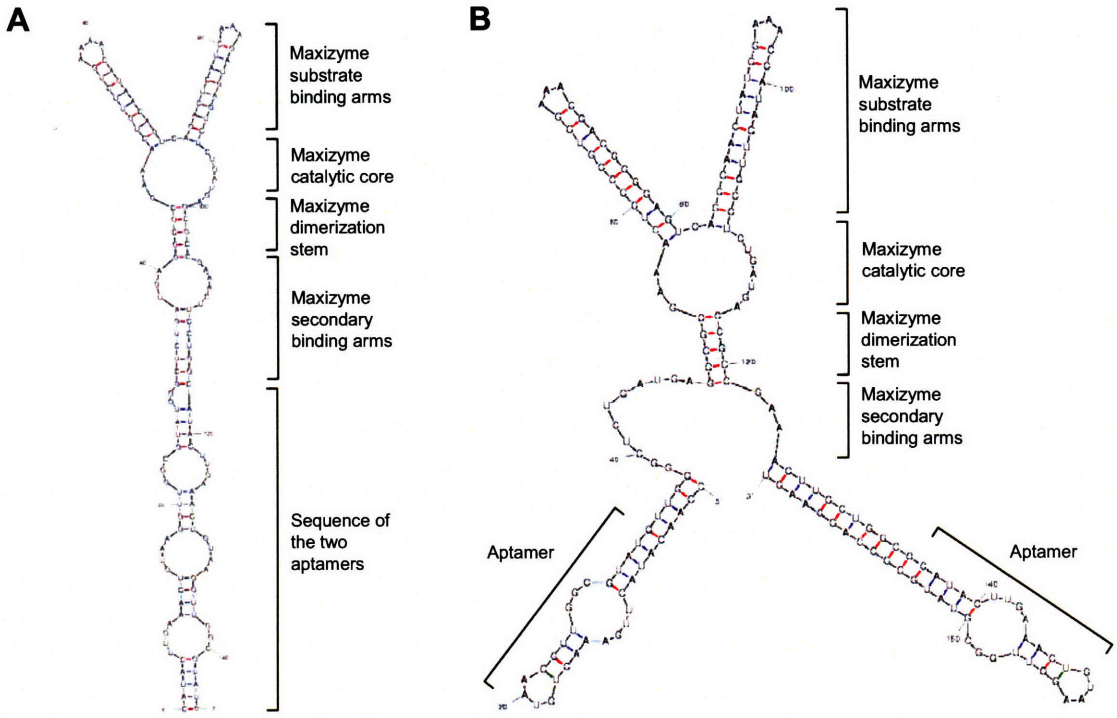


Figure 2-13: Secondary structures of the original (A) and optimized (B) aptazyme fusions as predicted by Mfold. Whereas the maxizyme domain is properly folded in both cases, the two aptamer domains only fold in a productive conformation in the optimized fusions. In the initial aptazyme fusions (A), the RNA aptamer fold is disrupted by interactions between the two aptamer domains and between the aptamer and the secondary binding arm of the left maxizyme monomer.

Aptazyme activity does not increase in the presence of a protein-protein interaction

NFkB p50 homodimerization affinity is lower than expected in the absence of DNA binding

We were then ready to examine whether the catalytic activity of the aptazyme fusion increased in the presence of a protein-protein interaction. Assuming a K_d for p50

homodimerization in the low nanomolar range, as it had been previously reported[52], we reasoned than incubating the aptazyme fusions in the presence of nanomolar concentrations of p50 would generate enough protein dimers to result in maxizyme activation. However, initial tests showed that the cleavage activity of the aptazyme fusion did not significantly change in the absence or presence of protein. Changing the aptazyme concentrations from 1–10 nM or the protein concentrations 1 nM–100 nM did not result in changes in ribozyme activity either (data not shown). We then realized that the reported nanomolar affinity had been estimated based on DNA gel shift assays, where the protein is incubated with its target DNA duplex, and protein dimerization is reported as the amount of protein with shifted mobility (i.e., DNA-bound) in a non-denaturing gel. As it had been proposed that the target DNA promotes protein homodimerization, the extent of dimerization in a gel-shift assay does not reflect the intrinsic protein dimerization affinity in the absence of DNA. Furthermore, we found that Ghosh and co-workers had previously estimated a low micromolar affinity for p50 homodimerization in the absence of DNA using analytical ultracentrifugation[53].

Unfortunately, with a working concentration of 10 nM aptazyme (this was the maximum concentration that showed no cleavage activity in the absence of protein, see Figure 2-11), incubation with micromolar concentrations of p50 would result in each aptazyme monomer binding to a different protein dimer, thus yielding no catalytically active aptazyme dimers. Based on this finding, we delineated two different approaches for testing our PPI detection strategy on the p50 homodimerization proof-of-principle:

1. Modifying the aptazyme fusion in order to increase its dimerization dissociation constant, such that we could work at micromolar concentrations of both aptazyme and protein.
2. Finding a way to promote p50 dimerization at nanomolar concentrations, such that we could work with the current aptazyme fusion.

We first focused on approach number one. It had been reported that decreasing the number of base pairs in the dimerization stem of the maxizyme resulted in higher dissociation constants from K_d(5bp) = 220 nM to K_d(3bp) = 550 nM or K_d(2bp) = 1 μM[27]. Although this report was promising, when we tested the constructs with two or three G-C base pairs in the dimerization stem, we did not observed any cleavage activity even at

concentrations significantly above their K_d (up to 10μM). We therefore decided to switch to the second approach.

Generating an obligate dimer of p50

We first considered utilizing the small protein myotrophin as a promoter of p50 homodimerization. Myotrophin is a 12kDa protein involved in NFκB-mediated cardiac hypertrophy[54]. Although its mechanism of action is still not clearly understood, it had been postulated that myotrophin acts as a chaperon to convert the transcriptional activators p50-p65 heterodimers into the transcriptional repressors p50-p50 and p65-p65 homodimers[52]. More importantly, contrary to what had been previously suggested[55], Chi-Wang Lin, a graduate student in our laboratory, demonstrated, using a novel crosslinking assay, that myotrophin could promote p50 homodimerization, at least partially, in the absence of p65 and DNA duplex[56]. We then tested the cleavage activity assay of the aptazyme fusions in the presence of varying concentrations of p50 and myotrophin but observed no activation. We hypothesized that the amount of protein dimers formed in the presence of myotrophin sufficed to be detected in a crosslinking assay but was not enough to result in detectable aptazyme activation.

It was clear at this point that we needed to find a way of artificially generating an obligate dimer of p50. We prepared obligate p50 dimers in two different ways: by chemical crosslinking of purified p50 protein and by recombinant expression of a gene encoding two p50 ORFs in tandem. In the following paragraphs, we will describe both approaches.

We first tested the most general crosslinking procedure, based on glutaraldehyde reaction with primary amines in the lysine side-chain of proteins. Glutaraldehyde crosslinking of p50 had been previously used to study the p50 dimer interface by Ghosh and co-workers, who reported optimal reaction conditions at 0.66 μM p50 and 1% w/v glutaraldehyde[53]. However, even at such low concentration of p50, our reaction mixtures contained a high percentage of un-reacted monomer (~ 50%) and higher-order oligomers (~ 10%) (data not shown). We then switched to cysteine crosslinking agents reasoning that, because there are only two cell-surface cysteines present in the p50 RHR[46], the crosslinking reaction should mainly yield monomer and dimer, with few higher-order

oligomers. As expected, crosslinking of p50 with either bis(maleimido) hexane (BMH) or 1,4-Di-[3′-(2′-pyridyldithio)-propionamido]butane (DPDPB) resulted in reaction mixtures that contained monomer and dimer almost exclusively. After optimizing reaction time and temperature, and protein and reagent concentrations (see Experimental section), BMH crosslinking resulted in a monomer to dimer ratio of 0.36 whereas DPDPB never resulted in a ratio higher than 0.6. Subsequent purification with Ni-NTA using gradient elution partially separated the monomer (one His_6 tag) from the dimer (two His_6 tags), ultimately decreasing the monomer to dimer ratio for the BMH reactions to ~ 0.2. We used this purified mixture of monomer and dimer in our aptazyme *in vitro* activity assays.

In parallel, we cloned a tandem sequence with two p50 ORFs in the pRSETb vector. Recombinant expression in BL21 cells using this pRSETb plasmid resulted in extremely low levels of expressed protein. Addition of varying concentrations of chemical chaperones such as betaine (2.5–25 mM), glycerol (0.5–5%), and DMSO (100 mM) did not detectably increase protein expression. We also tried co-expression with the protein chaperones dnaK, J and E, or GroES and GroEL with no success. Sub-cloning the tandem p50 sequence into a pET15b vector did however make a dramatic difference. The tandem p50 dimer expressed at very high levels in BL21 cells transformed with the $(p50)_2$-pET15b plasmid, although a small amount of p50 monomer was also observed in the lysates, probably due to tandem proteolysis. The monomer was separated from the dimer by two subsequent affinity purification steps: first, the lysate was purified using Ni-NTA agarose (the tandem carried an N-terminal His_6 tag); second, the eluate from the Ni-NTA column was further purified with Anti-FLAG-agarose (the tandem p50 had a C-terminal FLAG epitope). This purification scheme yielded the tandem p50 dimer with > 90% purity.

Aptazyme activity does not significantly increase in the presence of an artificial p50 dimer

We next tested the cleavage activity of the aptazyme fusions in the presence of the artificial p50 dimers. The initial results (Figure 2-14A) suggested that modest levels of activation were obtained in the presence of either the tandem dimer or the crosslinked

dimer. Quantification of several experiments like the one shown in Figure 2-14A never resulted in an activation factor higher than 1.2-fold, where the activation factor was calculated as the ratio of amount of product formed in the presence and in the absence of protein. Furthermore, similar activation factors were estimated for the maxizyme alone, in the absence of an aptamer domain (Figure 2-14B). It is worth noting that the observed aptazyme and maxizyme activations were both eliminated if the reaction was performed in the presence of excess α-p50 aptamer but not excess of an unrelated RNA sequence (data not shown), thus suggesting that activation was indeed caused by binding to the p50 dimer. However, varying the concentrations of aptazyme, maxizyme, and protein did not improve the ratio of aptazyme activation *versus* maxizyme activation (data not shown).

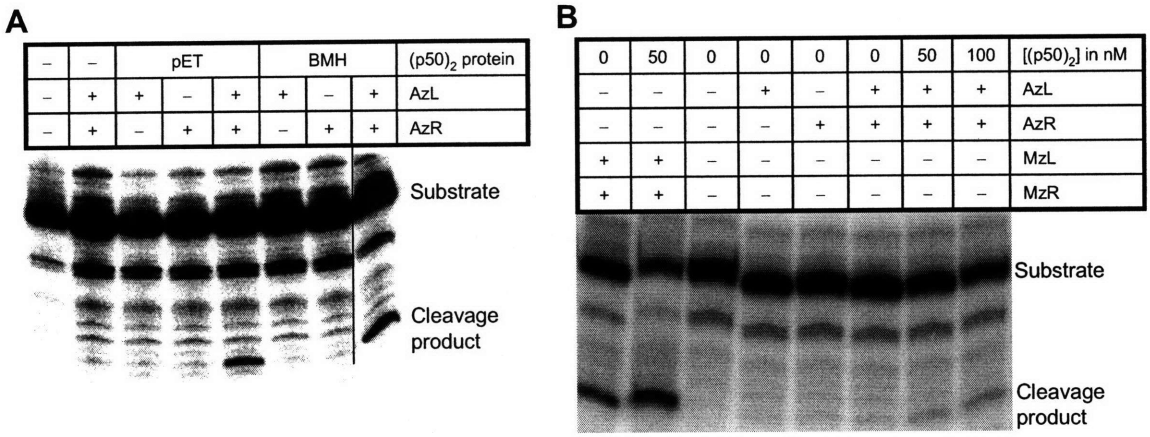


Figure 2-14: *In* vitro cleavage activity of the optimized aptazymes in the presence of p50 dimer. (**A**) The cleavage activity of the optimized aptazyme fusions was tested in the presence of 50 nM p50 dimer prepared by either recombinant expression of a tandem dimer (pET lanes) or by chemical crosslinking of the p50 monomers (BMH lanes). Negative controls are included without aptazyme (lane 1), without protein (lane 2) or in the presence of only one of the aptazyme monomers (lanes 3, 4, 6 and 7). (**B**) The activation of aptazyme cleavage activity (lanes 3–8) was tested side-by-side with that of the maxizyme constructs (lanes 1 and 2). The cleavage activity of the maxizyme was assayed in the absence (lane1) or presence (lane 2) of 50 nM p50 tandem dimer. The aptazyme was assayed in the absence of dimer (lane 6) or in the presence of 50 nm (lane 7) or 100 nM (lane 8) protein. Negative controls with no aptazyme (lane 3) or in the presence of only one of the monomers (lanes 4 and 5) were also included.

The maxizyme is active in cells as shown by a luciferase reporter

Finally, in parallel with the *in vitro* assays, the original maxizyme constructs targeting the S-19 substrate within the HIV-1 *tat* gene (Figure 2-2C) were tested for

in vivo activity. Following a published procedure[51], we transfected HeLa cells with a plasmid encoding a chimeric fusion of the long-terminal repeat (LTR) of HIV-1 and the luciferase gene. The LTR of HIV-1 contains regulatory elements that include a trans-activation responsive element region (TAR), which is a binding site for the HIV-1 regulatory protein Tat. Thus, expression of the Tat protein causes an increase in the transcription of the LTR-luciferase chimeric gene, ultimately resulting in an increase of luciferase enzyme expression levels. In contrast, expression of active maxizyme in the same cells should cause a decrease in the transcript levels of the *tat* gene, thus resulting in reduced luciferase activity. The results are shown in Figure 2-15. As expected, luciferase activity was maximal on cells expressing both the LTR-Luciferase chimeric gene and the Tat protein. Co-expression with the maxizyme constructs resulted in a ~ 5-fold reduction in luciferase activity, and this reduction required both maxizyme monomers. Although it is clear from Figure 2-15 that the maxizyme was active in live cells, the inhibition of luciferase activity that we observed in our experiments was 3.5-fold lower than that obtained by Taira and co-workers, who reported 90% inhibition using the same maxizyme constructs[51]. We also noticed that our results were hard to reproduce and were highly variable among samples (see large error bars in Figure 2-15).

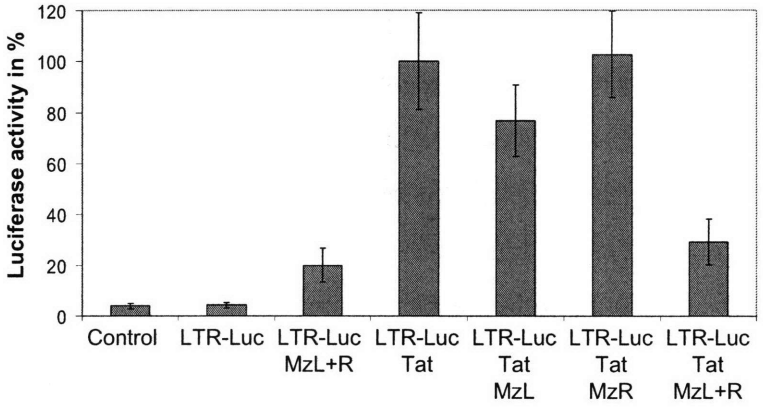


Figure 2-15: Maxizyme cleavage activity in mammalian cells. HeLa cells were transfected with plasmids for expression of the maxizyme constructs targeting the HIV-1 Tat mRNA (MzL and MzR), together with plasmids for expression of the Tat protein (Tat) and of the chimeric gene LTR-Luc, which expresses luciferase upon binding of the Tat protein. Negative controls that leave out one or more of the plasmids are also included. Luciferase activity was measured in lysates using the luciferin substrate. Luciferase activity for each sample is reported relative to that of cells expressing only the LTR-Luciferase and *tat* genes. Each data point represents the average of three independent experiments. Error bars, 1 s.d.

Discussion on the feasibility of the proposed detection scheme

While many of the problems encountered during the realization of this project are attributable to poor choice of the proof-of-principle system, we feel that they are mostly a reflection of the strict requirements of the proposed strategy. The suggested methodology involves four different binding processes, and, as such, it is governed by four different binding affinities. This is illustrated in Figure 2-16A. First, we require that the dissociation constant for dimerization of the aptazyme construct ($K_d^{AzL/R}$) be large enough such that, in the absence of a PPI, the majority of the aptazyme molecules are in monomeric (inactive) form. According to the *in vivo* results reported by Taira on co-workers, even for the dimeric maxizyme with only two base pairs in the dimerization stem (*in vitro* $K_d^{MzL/R} \sim 1$ μM), the percentage of maxizyme molecules present in dimeric form is enough to result in ~ 90% inactivation of the gene of interest in mammalian cells. We must realize that this reported *in vivo* activity corresponds in our system to the background activity in the absence of a PPI. Fortunately, the target reporter gene in our system (e.g., β-lactamase) is also transfected and thus present at much higher concentrations than those of the endogenous genes in Taira's reports. This means that, perhaps, a higher percentage of dimeric aptazyme would be required for complete inactivation (i.e., although, owing to the signal amplification of the ribozyme catalytic activity, this may not be the case). It therefore remains unknown what the actual percentage of active cellular dimeric aptazyme is, and whether introduction of a PPI-mediated aptazyme dimerization would result in a large enough increase in activity so as to generate a significant signal-to-background ratio (i.e., in the best protein complementation assays, the PPI introduces a ~ 10-fold increase in reporter activity). This will ultimately depend of both, the $K_d^{AzL/R}$ and the actual concentrations of aptazyme achieved in cells; however, since higher aptazyme concentrations are desired in order to cover all the protein binding sites, a high $K_d^{AzL/R}$ is needed.

Second, the dissociation constant of the PPI ($K_d^{(p50)_2}$, or its equivalent for a different protein pair K_d^{PPI}) must be higher than that of aptazyme dimerization, to ensure that the PPI increases the percentage of aptazyme molecules that are in dimeric form. Third, the affinity of the aptamer should be characterized by a dissociation constant ($K_d^{Az\text{-}Prot}$) low enough to guarantee that, even at the lowest possible aptazyme

concentration, the majority of the aptazyme molecules are bound to the target protein. RNA aptamers have been previously selected with picomolar dissociation constants[57] and thus, this requirement is more easily achieved than the other K_d requirements.

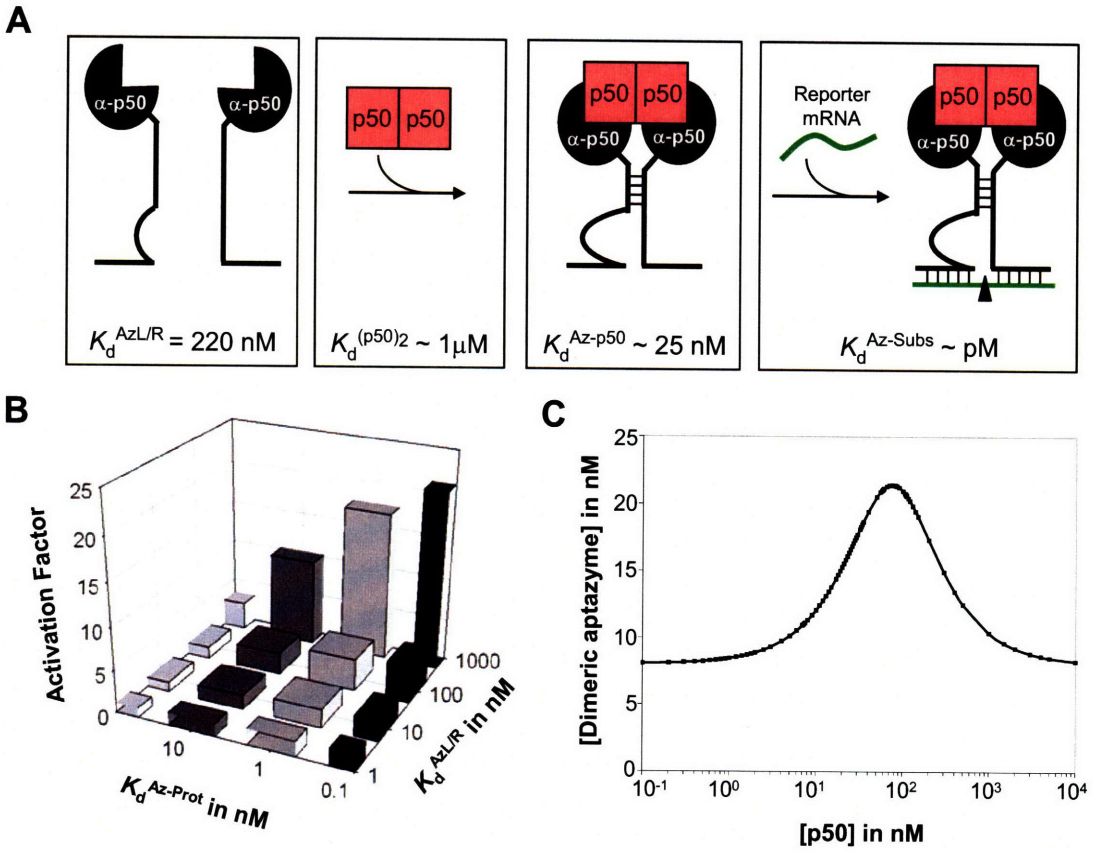


Figure 2-16: Requirements for the RNA-based detection of endogenous protein-protein interactions. (**A**) Description of the four binding events involved in the detection and the dissociation constants that govern them. $K_d^{AzL/R}$: dissociation constant for aptazyme dimerization. $K_d^{(p50)2}$: dissociation constant for p50 homodimerization. K_d^{Az-p50} dissociation constant for aptazyme binding to p50. $K_d^{Az-Subs}$: dissociation constant for aptazyme binding to its mRNA substrate. (**B**) Simulation of activation factor values for several combinations of $K_d^{Az-Prot}$ and $K_d^{AzL/R}$. The activation factor is calculated as the ratio of the amount of dimeric aptazyme in the presence and absence of a PPI. The corresponding amounts of dimeric aptazyme were calculated from the binding equation described in the Experimental section. The total aptazyme concentration was assumed to be 50 nM in all cases. The total protein concentrations were 50, 60, and 150 nM for $K_d^{Az-Prot}$ 0.1 and 1, 10, and 100 respectively; which were calculated as optimal (see panel C in this figure). (**C**) Dependence of the percentage of dimeric activated aptazyme molecules on the concentration of target proteins. This graph was obtained for a total aptazyme concentration of 50 nM and the dissociation constants obtained for our optimized aptazyme constructs, $K_d^{AzL/R}$ = 220 nM and $K_d^{Az-Prot}$ = 25 nM.

Ultimately, for this method to be useful, we would need to maximize the activation factor, that is, the ratio of dimeric aptazyme in the presence *versus* the absence of a PPI. Figure 2-16B simulates the values of this activation factor at different combinations of $K_d^{Az\text{-}Prot}$ and $K_d^{AzL/R}$. This simulation quantifies the previous discussion: highest activation factors are obtained when $K_d^{Az\text{-}Prot}$ is lowest and $K_d^{AzL/R}$ is highest. It must also be noted that, assuming an aptazyme concentration in cells of 50 nM (used for the simulation in Figure 2-16B) the threshold activation factor of 10 is only achieved at $K_d^{AzL/R} \sim 1$ μM. If the cellular concentrations of aptazymes are higher, so must the $K_d^{AzL/R}$ be. Finally, Figure 2-16C illustrates the fact that the activation factor varies greatly with the concentration of target protein. This means that, for a defined set of K_d values, relatively small changes in protein concentration can result in different aptazyme activities, thus resulting in false positives.

The preceding discussion on interaction affinities does however not explain our failure to detect an increase in aptazyme activity in the presence of the obligate p50 dimer. In this context, the effective concentration of p50 is such that it resembles a PPI with infinitely low K_d^{PPI}. This failure thus illustrates yet another constrain of the system: the geometric requirement. We hypothesize that we did not observe aptazyme activation because the two aptazyme monomers were not brought into close proximity upon binding to the p50 obligate dimer. This may have been caused by a short linker in the case of the tandem dimer or by a p50-p50 interaction geometry created during cysteine crosslinking that is different from the one revealed by the $(p50)_2$-RNA crystal structure. This requirement adds another layer of complexity to the already complex aptamer selection process. Selection of new aptamers to be used in this methodology would require the inclusion of the necessary steps to ensure that:

1. Aptamer binding does not disrupt the PPI (i.e., it uses a different interaction surface). This could be done by performing the selection in the presence of the protein partner.
2. Aptamer binding is not invasive; that is, it does not affect the biological activity of the target protein. This was not the case in our proof-of-principle, in which the RNA aptamer binds to the DNA-binding interface. This requirement may be easily fulfilled in the case of transcription factors, where

the selection could be performed for example in the presence of the target DNA. It would however become more complicated for other proteins for which many PPIs are not well characterized or even unknown.

3. Aptamer binding to both protein partners brings the aptazyme monomers into close proximity. The ability to find this aptamer would require the *a priori* knowledge of how the target proteins interact, and/or a sophisticated SELEX strategy that incorporated a readout of aptazyme activity.

Furthermore, *in vivo* genetic selections may be needed post-SELEX to guarantee that the aptamer retains binding activity when expressed in eukaryotic cells[58].

Finally, one of the biggest question marks for this system is whether it has a real potential to work inside living cells. We already emphasized in the introduction to this chapter that the relatively simple and repetitive composition of RNA molecules makes them easy to synthesize, allows for selection of improved characteristics, and makes it possible to manipulate them *in vitro*. Problems arise, however, when expressing exogenous RNAs *in vivo*, in ensuring correct intracellular localization, providing protection against nucleolytic attack, ensuring access to target sites, and achieving similar cleavage activity to that observed *in vivo*. Work in the Taira lab during the past 10 years has addressed most of these problems for the maxizyme system. They created hybrid ribozymes that combine the cleavage activity of the maxizyme with the cytosolic export activity of a $tRNA^{Val}$ promoter and the unwinding activity of an endogenous RNA helicase, and showed that these were capable of gene inactivation in a variety of cell-lines (these efforts have been recently reviewed by Sano and Taira)[59]. However, what has not yet been addressed is the maxizyme need for high Mg^{2+} concentrations.

Although there have been reports by the Taira group and others of the minimal hammerhead ribozyme and its maxizyme derivative functioning *in vivo*[20,21], there is also evidence that the minimal hammerhead ribozyme is not optimal for *in vivo* function. In 2003, a report provided insight into why natural HHRzs function effectively in the cellular environments whereas artificial ones do not. Khvorova and co-workers showed that the non-conserved nucleotides, which had been deleted from the natural Rzs to obtain the minimal HHRz (and its derivative, the maxizyme), are in fact responsible for adoption of the appropriate conformation for cleavage[60]. The interaction between the

loops of stem I and stem II (Figure 2-17B), through two Watson-Crick base pairs and some additional non-Watson-Crick interactions, is required for the correct folding and the strong activity of the ribozyme at physiological concentrations of Mg^{2+} (which are ~ 100 fold lower than the concentrations commonly used in *in vitro* assays)[61]. Both loops are absent in the trans-cleaving conventional HHRz (Figure 2-17A). It is now believed that the tertiary interaction stabilizes the catalytic core in the active conformation, with the metal ion binding site pre-organized, and thus allowing better metal binding[62].

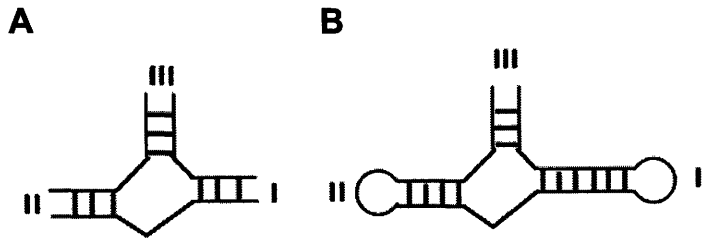


Figure 2-17: Comparison of secondary structures of the minimal hammerhead ribozyme (A) and the natural hammerhead ribozyme from tobacco ringspot satellite virus (B). Figure adapted from Khvorova et al[60].

This important finding suggests that the loops of stem I and II should therefore be incorporated in the design of any ribozyme designed to work inside cells. It is however hard to imagine how this could be done for the maxizyme, whose stem II has been even further minimized to serve as the dimerization stem. A completely different ribozyme platform may thus be needed if we were to design a PPI detection platform with maximal *in vivo* activity.

Conclusions

In this chapter we have presented our efforts to develop a new methodology to detect endogenous protein-protein interactions in live cells. The method is an adaptation of the classical protein complementation assays, where each half of a reporter protein is fused to the proteins of interest. In our proposed methodology, we use an RNA aptamer as a scaffold to detect endogenous proteins, and a dimeric ribozyme as the enzyme to be complemented upon protein-protein interaction. We call the fusion of the RNA aptamer

to the dimeric ribozyme an aptazyme. We have described our engineering process to develop aptazyme fusions where the two domains (aptamer and ribozyme) can fold and function independently of each other, and we have shown that the engineered aptazymes are catalytically active *in vitro*, in the absence of protein, at concentrations higher than 10 nM. The aptazyme fusions can also be modulated to recognize and cleave different transcripts, such as the mRNA of the β-lactamase enzyme. Additionally, we have shown that different cleavage sites within the β-lactamase mRNA are more accessible than others, probably because they lie within loops or unstructured regions of the mRNA. We also characterized the aptazyme fusions for their binding affinity to the p50 protein and found their equilibrium dissociation constants (32.5 ± 5.1 for AzL and 21.5 ± 2.4 for AzR). Finally, we have described our efforts to modulate the aptazyme so that its catalytic activity increases in the presence of a p50 dimer caused by PPI-induced aptazyme dimerization. Unfortunately, no such activation was ever observed *in vitro* under many different assay conditions. We hypothesized that the lack of activation was caused by incorrect geometry of the p50 dimer, which did not bring the two ribozyme monomers into close enough proximity to allow for dimerization. This hypothesis illustrates the geometric requirement of the proposed system. Other system constrains include the need for strict control of the interaction affinities between the three components of the system (aptazyme-aptazyme, aptazyme-protein, and PPI), and the need for selection of aptamers that do not disrupt the biologically-relevant PPIs of the target protein. Additionally, the recent finding that optimal *in vivo* ribozyme activity requires several nucleotide sequences that are missing in our aptazyme further limits the applicability of this methodology. We propose that, if a similar approach is attempted in the future, a ribozyme platform different that the maxizyme be used.

Experimental

RNA *in vitro* transcription using T7 RNA polymerase

The substrate, aptamer, ribozyme, maxizyme, and aptazyme RNA transcripts were generated by *in vitro* transcription using the Ampliscribe T7-*Flash* Transcription kit (Epicentre Biotechnologies) following the manufacturer's protocol. DNA templates were chemically synthesized by MWG, and were added to the transcription reaction without further purification. Table 2-3 summarizes the DNA templates employed in the transcription of all the constructs mentioned in the text.

Table 2-3: Oligos used in the *in vitro* transcription reactions. Nucleotides in italic are needed for transcription. These anneal to the primer 5′ GCGTAATACGACTCA CTATAG 3′ to form the double-stranded T7 promoter that the T7 RNA polymerase recognizes to start transcription. Highlighted in grey are the nucleotides corresponding to either the substrate or the ribozyme substrate-binding arms. Nucleotides in blue represent the sequence of the RNA aptamer against the p50 protein. Of those, the nucleotides in bold highlight the point mutations that disrupt protein binding; mutant 1 includes a U11:A20 to C11:G20 mutation, mutant 2 includes a G12:C19 to A12:U19 mutation, and mutant 3 includes both mutations. Nucleotides in red represent the catalytic core of the ribozyme or maxizyme. Nucleotides in bold represent the dimerization stem of the maxizyme. Nucleotides in green represent nucleotides added to promote independent folding of the aptamer and ribozyme domains in the aptazyme fusions. For simplicity, the oligos used for transcribing the non-binding point mutants of the aptazymes have not been included in this table. To generate these oligos, please refer to the mutations highlighted in the oligos used to transcribe the aptamer. The asterisk (*) denotes maxizyme constructs with two base pairs in the dimerization stem instead of five.

Construct		Transcription Oligo (5′ to 3′)
Substrate	S19	*GG*GATGAGTCTGACTGTTCTG*CCTATAGTGAGTCGTATTACGC*
	Bla-1	*GG*AATAAGGGCGACACGGAAA*CCTATAGTGAGTCGTATTACGC*
	Bla-2	*GG*GTATGCGGCGACCGAGTTG*CCTATAGTGAGTCGTATTACGC*
	Bla-3	*GG*AGTCATTCTGAGAATAGTG*CCTATAGTGAGTCGTATTACGC*
	Bla-4	*GG*GCGTTTCTGGGTGAGCAAAAACAGG*CCTATAGTGAGTCGTATTA CGC*
	Bla-5	*GG*GGGAGGGCTTACCATCTGG*CCTATAGTGAGTCGTATTACGC*
	Bla-6	*GG*TAGTTGCCTGACTCCCCGT*CCTATAGTGAGTCGTATTACGC*
	Bla-6L	*GG*CCATAGTTGCCTGACTCCCCGTCG*CCTATAGTGAGTCGTATTAC GC*

Aptamer	WT		GGCATACGCCAACCTTACAGTTTCAAGTATGCCTATAGTGAGTCGTATTACGC
	Mut.1		GGCATACGCCAGCCTTACAGCTTCAAGTATGCCTATAGTGAGTCGTATTACGC
	Mut.2		GGCATACGCCAATCTTACAATTTCAAGTATGCCTATAGTGAGTCGTATTACGC
	Mut.3		GGCATACGCCAGTCTTACAACTTCAAGTATGCCTATAGTGAGTCGTATTACGC
HHRz	Bla-1		GGTTTCCGTGTTTCGGCGTTTCGGCCTCATCAGGCCCTTATTCCTATAGTGAGTCGTATTACGC
	Bla-2		GGCAACTCGGTTTCGGCGTTTCGGCCTCATCAGGCCGCATACCCTATAGTGAGTCGTATTACGC
	Bla-3		GGCACTATTCTTTCGGCGTTTCGGCCTCATCAGAGAATGACTCCTATAGTGAGTCGTATTACGC
Maxizyme	Left	S-19	GGCCAGGAAGTTTCGGCGCTCATCAGAGACTCATCCCTATAGTGAGTCGTATTACGC
		S-19*	GGCCAGGAAGTTTCGCTCATCAGAGACTCATCCCTATAGTGAGTCGTATTACGC
		Bla1	GGCCAGGAGACTTCGGCGCTCATCAGGCCCTTATTCCTATAGTGAGTCGTATTACGC
		Bla4	GGCCAGGAGACTTCGGCGCTCATCAGACCCAGAAACGCCCTATAGTGAGTCGTATTACGC
		Bla5	GGCCAGGAGACTTCGGCGCTCATCAGAGCCCTCCCCCTATAGTGAGTCGTATTACGC
		Bla6	GGCCAGGAGACTTCGGCGCTCATCAGAGGCAACTATGGCCTATAGTGAGTCGTATTACGC
	Right	S-19	GGCAGAACAGTTTCGCGCCTCATCAGAGCCCCTATAGTGAGTCGTATTACGC
		S-19*	GGCAGAACAGTTTCGCTCATCAGAGCCCCTATAGTGAGTCGTATTACGC
		Bla1	GGTTTCCGTGTTTCGCGCCTCTGTAGACTTCCTATAGTGAGTCGTATTACGC
		Bla4	GGCCTGTTTTTGCTTTCGCGCCTCTGTAGACTTCCTATAGTGAGTCGTATTACGC
		Bla5	GGCCAGATGGTTTCGCGCCTCTGTAGACTTCCTATAGTGAGTCGTATTACGC
		Bla6	GGCGACGGGGAGTTTCGCGCCTCTGTAGACTTCCTATAGTGAGTCGTATTACGC
Aptazyme-WT	Left	0	GGCATACGCCAACCTTACAGTTTCAAGTATGCCAGGAAGTTTC GGCGCTCATCAGAGACTCATCCCTATAGTGAGTCGTATTACGC
		1	GGACTTCCTGGCATACGCCAACCTTACAGTTTCAAGTATGCCAGGAAGTTTC GGCGCTCATCAGAGACTCATCCCTATAGTGAGTCGTATTACGC
		2	GGACTTCCTGGTTGGCATACGCCAACCTTACAGTTTCAAGTATGCCTTCCAGGAAGTTTC GGCGCTCATCAGAGACTCATCCCTATAGTGAGTCGTATTACGC
		2.1	GGGTCTCCTGGTTGGCATACGCCAATCTTACAATTTCAAGTATGCCTTCCAGGAGACTTC GGCGCTCATCAGAGGCAACTATGGCCTATAGTGAGTCGTATTACGC
		2.1*	GGGTCTCCTGGTTGGCATACGCCAATCTTACAATTTCAAGTATGCCTTCCAGGAGACTTC GCTCATCAGAGGCAACTATGGCCTATAGTGAGTCGTATTACGC
		3	GGGGCATACGCCAACCTTACAGTTTCAAGTATGCCTTCCAGGAAGTTTC GGCGCTCATCAGAGACTCATCCCTATAGTGAGTCGTATTACGC
		4	GGGGAACATACGCCAACCTTACAGTTTCAAGTATGTTCCTTCCAGGAAGTTTC GGCGCTCATCAGAGACTCATCCCTATAGTGAGTCGTATTACGC

Aptazyme-WT	Right	0	GGCAGAACAGTTTCGCGCCTCATCAGAGCCCATACGCCAACCTTACAGTTTCAAGTATGCCTAT AGTGAGTCGTATTACGC
		1	GGCAGAACAGTTTCGCGCCTCATCAGAGCCGGCATACGCCAACCTTACAGTTTCAAGTATGCCC CTATAGTGAGTCGTATTACGC
		2	GGCAGAACAGTTTCGCGCCTCATCAGAGCCGGAACATACGCCAACCTTACAGTTTCAAGTATGT TCCCCTATAGTGAGTCGTATTACGC
		2.1	GGCGACGGGGAGTTTCGCGCCTCTGTAGACTTGGAACATACGCCAATCTTACAATTTCAAGTAT GTTCCCCTATAGTGAGTCGTATTACGC
		2.1*	GGCGACGGGGAGTTTCGCTCTGTAGACTTGGAACATACGCCAATCTTACAATTTCAAGTATGTT CCCCTATAGTGAGTCGTATTACGC

Before initiating the transcription reaction, 5 µg of DNA template were combined with an equimolar amount of the T7 promoter complementary oligo (Table 2-3, legend) and the mixture was heated to 95 °C and allowed to slowly cool to room temperature to promote annealing of the two oligos. Thereafter, the remaining components of the transcription reactions were added: 7.2 µl of each rNTP (final concentration 9 mM), 10 µl of the 10x reaction buffer provided by the manufacturer, 10 µl T7 RNA polymerase enzyme, and RNAse-free water to adjust the total volume to 100 µl. The transcription reaction was allowed to proceed at 42 °C for 30 minutes–2 hours, until a white precipitated appeared (insoluble complex between Mg^{2+} ions and inorganic phosphate). Purification was done by PAGE (10–20% acrylamide) followed by crush and soak RNA extraction from the acrylamide, and ethanol precipitation. RNA concentration was estimated by measuring A_{260} using an extinction coefficient of 40 $mL.\mu g^{-1}cm^{-1}$.

Removal of 5′ tri-phosphate using calf intestinal phosphatase (CIP)

50 pmoles of RNA were combined with 10 units of CIP enzyme (New England Biolabs) and phosphatase buffer (New England Biolabs) in a total volume of 10 µL. After incubation at 37 °C for 1 hour, the reaction was purified by phenol-chloroform extraction followed by ethanol precipitation. To perform the extractions, the reaction volume was increased to 50 µL and NaCl was added to $C_f = 0.3$ M. Three extractions with 50 µL of phenol: chloroform 3:1 were followed by one extraction with an equal volume of

chloroform (for each extraction, the mixture was vortexed for 2 minutes followed by centrifugation for 2 minutes). The RNA was then precipitated by addition of 1 μg of glycogen and three volumes of ethanol, incubation at – 80 °C for 30 minutes, and centrifugation at 13,200 rpm for 30 minutes at 4 °C. Thereafter, the resulting pellet was washed once with cold 70% ethanol, allowed to air-dry, and resuspended in 8 μL of RNAse-free water.

RNA radioactive end-labeling

De-phosphorylated RNA transcripts were 5′ labeled with ^{32}P using T4 polynucleotide kinase (New England Biolabs). 50 pmoles of RNA were combined with DTT (C_f = 10 mM), 20 units of kinase enzyme, 0.25 μl of [^{32}P-γ] ATP (GE Healthcare, 10 μCi/μl), and kinase buffer (New England Biolabs) in a total volume of 50 μL, and incubated at 37 °C for 1 hour. Unincorporated [^{32}P-γ] ATP was removed by gel filtration using Sephadex G25 resin (QuickSpin columns, Roche).

***In vitro* ribozyme activity assay**

All ribozyme activity assays were performed under single turnover conditions. The reaction conditions were as follows: 1 nM end-labeled substrate (5000cpm), 10 nM–1 μM RNA enzyme (HHRz, Mz or Az), 100 mM NaCl, 1 mM DTT, and 10 mM $MgCl_2$, in 10 mM HEPES buffer pH 7.5 or 50 mM Tris pH 8. In some cases, 0.25 μL RNAsin (RNase inhibitor, Promega) was included in the mixture. The reactions were first heated to 95 °C for 2 minutes in the absence of $MgCl_2$, and allowed to cool to 37 °C to facilitate annealing of the RNA enzyme to its substrate. Thereafter, the reactions were initiated by addition of 10 mM $MgCl_2$. In the cases where the activity of the ribozyme was assayed in the presence of p50 protein, the mixtures were allowed to cool down to room temperature before adding the protein, and then incubated at room temperature for 10 minutes to allow binding of the aptamer before initiating the reactions with 10 mM $MgCl_2$. After incubation for 5 minutes – 1 hour at 37 °C the reactions were quenched by addition of an equal volume of gel loading buffer (8 M urea, 25 mM EDTA, 0.25%

bromophenol blue, 0.25% xylene cyanol, 40% sucrose). The reaction mixtures were separated on a denaturing 20% acrylamide gel (8 M urea) and the gels were exposed to phosphor-screens (GE Healthcare) overnight at –20 °C. Autoradiograms were visualized with a Storm 860 instrument (GE Healthcare).

Bacterial expression and purification of the human p50 protein

The plasmid for expression of p50 was a gift from Prof. Ghosh (UC San Diego) and it encodes an untagged truncated murine p50 (amino acids residues 39 to 363). Recombinant expression and purification were essentially done as previously reported by Lebruska et al[39]. Briefly, the plasmid was transformed into *E. coli* BL21 (DE3) cells, which were amplified in LB media supplemented with 100 μg/mL ampicillin at 37 °C until OD_{600} 0.2. Protein expression was induced with 0.2 mM isopropyl-β-D-thiogalactopyranoside (IPTG) for approximately 3 hours, until cells reached OD_{600} 0.9. Thereafter, cells were harvested by centrifugation (6,000 rpm, 10 minutes, 4 °C) and the pellet was resuspended in lysis buffer (50 mM NaCl, 2.5 mM PMSF, and protease inhibitor cocktail in 50 mM Tris, pH 7.5). Cells were lysed by ultrasonic treatment (six 15-second bursts, with 1 minute of cooling to 4 °C between bursts). The extract was cleared by centrifugation (17,700 g, 10 minutes, 4 °C) and loaded onto a cation exchange column (22 mL of Q-Sephadex resin, GE Healthcare). Elution from the Q-Sephadex column was performed with two column volumes of a low salt buffer (50 mM NaCl, 50 mM Tris, pH 7.5) and the eluate was immediately loaded onto an anion exchange column (60 mL of SP Sepharose Fast Flow resin, GE Healthcare). The protein was then eluted at 1mL/minute using a 50–400 mM NaCl gradient over ten column volumes. Fractions were analyzed by 12% SDS-PAGE followed by Coomassie staining. Fractions containing mutant p50 (approx. fractions eluting at 160–190 mM NaCl) were pooled and dialyzed against aptamer binding buffer (100 mM NaCl and 1 mM DTT in 10 mM HEPES, pH 7.5). Protein concentrations were measured using the BCA assay (Pierce) with BSA as the reference standard.

Bacterial expression and purification of a non-dimerizing mutant of the human p50 protein

The plasmid for expression of the non-dimerizing mutant of p50 was a gift from Prof. Ghosh (UC San Diego) and encoded an N-terminally His_6-tagged truncated p50 (amino acid residues 39–363) with Y267D and L269D point mutations[63]. The plasmid was transformed into *E. coli* BL21(DE3) cells, which were amplified in LB media supplemented with 100 μg/mL ampicillin at 37 °C until OD_{600} 0.9. Protein expression was induced with 200 μg/mL isopropyl-β-D-thiogalactopyranoside (IPTG) at room temperature overnight. Thereafter, cells were harvested by centrifugation (6,000 rpm, 15 minutes, 4 °C) and the pellet was resuspended in lysis buffer (50 mM Tris base, 300 mM NaCl, pH 7.8) containing 2.5 mM phenylmethylsulfonyl fluoride (PMSF) and protease inhibitor cocktail (Calbiochem). Cells were lysed by ultrasonic treatment (six 30-second bursts, with 1 minute of cooling to 4 °C between bursts). The extract was cleared by centrifugation (21,000 rpm, 15 minutes, 4 °C) and the His_6-tagged protein was purified using Ni-NTA agarose (Qiagen). Fractions were analyzed by 12% SDS-PAGE followed by Coomassie staining. Fractions containing mutant p50 were pooled and dialyzed against aptamer binding buffer (100 mM NaCl and 1 mM DTT in 10 mM HEPES pH 7.5). Protein concentrations were measured using the BCA assay (Pierce) with BSA as the reference standard.

***In vitro* aptamer binding assay**

To determine the fractional saturation of RNA binding to p50, the reaction conditions were as follows: 0.9 nM end-labeled RNA aptamer or aptazyme fusion (5000 cpm) and 50 nM p50 protein in binding buffer (100 mM NaCl, and 1 mM DTT in 10 mM HEPES pH 7.5) supplemented with 50 μg/mL BSA. After incubation at room temperature for 30 minutes, the reactions were filtered over nitrocellulose membranes, which had been previously treated with binding buffer. The filters were then washed twice with ice-cold binding buffer, and air-dried. All the filtering and washing steps were performed using a vacuum manifold (Millipore). The radioactivity spotted on the filters

was measured by liquid scintillation counting. Negative control reactions performed in the absence of p50 were used for background correction.

Determination of equilibrium dissociation constants for p50 binding to the aptazyme fusions

The determination of the equilibrium dissociation constants was performed as previously reported by Lebruska et al[39]. Briefly, individual assays were performed as described above for the *in vitro* aptamer binding assay except that varying amounts of protein were used. The binding isotherms were then plotted and fitted to the equation:

$$\theta = \frac{1}{2[T]_t}\left[[T]_t + K_d + [P]_t - \sqrt{([T]_t + K_d + [P]_t)^2 - 4[P]_t[T]_t} \right]$$

where θ is the fractional saturation of the radiolabeled probe, T represents the RNA probe, P the p50 protein, and the subscript (t) refers to total concentration, and it need not be assumed that $[P]_t \sim [P]_{free}$. Curve fitting was performed using Sigma Plot software (Systat Software Inc.).

Bacterial expression and purification of myotrophin

The plasmid for expression of myotrophin was a gift from Prof. Sivasubramanian (Baylor College of Medicine, Houston, TX) and encoded the gene for human myotrophin within a pET3a-51 vector. Recombinant expression and purification were essentially done as previously reported by Sivasubramanian et al[54]. Briefly, the plasmid was transformed into *E. coli* BL21(DE3) cells, which were amplified in LB media supplemented with 100 μg/mL ampicillin at 37 °C until OD_{600} 0.2. Protein expression was induced with 0.1 mM IPTG at room temperature overnight. Thereafter, cells were harvested by centrifugation (6,000 rpm, 15 minutes, 4 °C) and the pellet was resuspended in lysis buffer (50 mM Tris base, 75 mM NaCl, pH 8) containing 2.5 mM PMSF and protease inhibitor cocktail (Calbiochem). Cells were lysed by freeze-thawing three times and the extract was cleared by centrifugation (21,000 rpm, 15 minutes, 4 °C). Myotrophin was highly overexpressed in the supernatant and was purified by subsequent separation steps using first a Centriprep-30 cartridge (30 kD cutoff, Amicon) and then a Centriprep-

10 (10 kD cutoff, Amicon). Both Centriprep columns were pretreated with 0.1% diethylpyrocarbonate (DEPC) for 2 hours at 37 °C to remove RNAse. Eluate and concentrate fractions of both separation steps were analyzed by 19% SDS-PAGE followed by Coomassie staining. Protein concentration of the concentrate fraction of the Centriprep-10 column was measured using the BCA assay (Pierce) with BSA as the reference standard.

Chemical crosslinking of p50 to generate p50 dimers

Glutaraldehyde crosslinking. Purified p50 (0.6 μM)was combined with 0.5% w/v glutaraldehyde for 15 minutes at 4 °C, followed by 20 minutes of additional incubation in the presence of 10 mg/ml of sodium borohydride. The reaction mixtures were analyzed by 10% SDS-PAGE followed by Coomassie or silver staining.

Cysteine crosslinking. Purified p50 was first extensively dialyzed against PBS pH 7.4 to remove the dithiothreitol (DTT) present in the storage buffer. The crosslinking reaction was then performed by incubation for 30 minutes at 30 °C with a 5-fold molar excess of either bis(maleimido) hexane (BMH, Pierce) or 1,4-Di-[3′-(2′-pyridyldithio)-propionamido]butane (DPDPB, Pierce). Starting concentrations of p50 ranged from 3.5–14 μM. The reaction was quenched by addition of 5 mM DTT.

Construction of the p50-tandem gene for expression of a p50 tandem dimer

The p50 sequence was introduced twice in tandem into the pRSETb vector (Invitrogen). The first p50 sequence was inserted between *BamHI and EcoRI u*sing the primers 5'CAT GAT AA*G GAT CC*G GGC CCA TAC CTT CAA ATA TT and 5′ CAT CAT *GAA TTC* GCC CGA GCC TTC GCC CGA GCC CGG TTT GCC CGA GCC CGA GGT CGA GCC CGA GCC CTT CTG GCG TTT CCT TTG CAC. The second p50 sequence was inserted between *EcoRI* and *HindIII* using the primers 5'GAT AAG *GAA TTC* TCG ACC TCG GGC TCG GGC AAA CCG GGC TCG GGC GAA GGC TCG GGC TCG GGC GGC CCA TAC CTT CAA ATA TT and 5' GAA TTC TC*A AGC TT*C **TTT GTC GTC GTC GTC TTT GTA GTC** CTT CTG GCG TTT CCT TTG CAC. The former primers also introduce a C-terminal FLAG epitope (in bold) and a 35-

amino acid (GSGSTSGSGKPGSGEGSGSTSGSGKPGSGEGSGSG) protease-resistant linker in between the two p50 proteins to allow for proper folding and dimerization (underlined). To insert the p50 tandem gene into pET15b, the p50-linker-p50-FLAG sequence was PCR-amplified from pRSETb using the primers 5' CAT CCC *CAT ATG* GGC CCA TAC CTT CAA ATA TTA G and 5' CAT CCC *GGA TCC* TCG AGT TTG TCG TCG TCG TCT TTG TAG and ligated into the NdeI and BamHI sites of pET15b (Novagen) to give $(p50)_2$-pET15b.

Bacterial expression and purification of the p50-tandem dimer

E. coli BL21(DE3) cells transformed with $(p50)_2$-pRSETb or $(p50)_2$-pET15b expression plasmids were amplified in LB media supplemented with 100 μg/mL ampicillin at 37 °C until OD_{600} 0.6. Protein expression was induced with 100 μg/mL IPTG for 3 hours at 30 °C. Cells were harvested by centrifugation and the pellet was resuspended in lysis buffer (50 mM Tris base, 300 mM NaCl, pH 7.8) containing 2.5 mM PMSF and protease inhibitor cocktail (Calbiochem). Cells were lysed by ultrasonic treatment (six 30-second bursts, with 1 minute of cooling to 4 °C between bursts). The extract was cleared by centrifugation (17,700 g, 10 minutes, 4 °C) and the N-terminally His_6-tagged protein was purified using Ni-NTA agarose (Qiagen). Fractions were analyzed by 12–16% SDS-PAGE followed by Coomassie staining and the fractions containing protein were pooled and dialyzed against TBS (50 mM Tris, 150 mM NaCl, pH 7.4). Thereafter, the protein mixture (p50 dimer plus monomer) was further purified using FLAG-M2-agarose (Sigma). After elution with five column volumes of 1 mM FLAG peptide, the purified dimer was dialyzed against aptamer binding buffer (100 mM NaCl and 1 mM DTT in 10 mM HEPES pH 7.5). Protein concentrations were measured using the BCA assay (Pierce) with BSA as the reference standard.

Live cell maxizyme activity assay

The plasmids for expression of the LTR-luciferase chimeric gene (pHyg-LTR-Luc)[64], the Tat protein (pCD-SRαTat)[51], and the maxizyme constructs (p5L, p5R, and pD5L/R)[51] were a gift of Prof. Taira (Univ. of Tsukuba). HeLa cells were transfected

with either one or a combination of the above plasmids, and lysates were generated 48 hours after transfection. Luciferase activity in the lysates was measured using the Luciferase Assay System (Promega) according to the manufacturer's instructions.

References

1. Galarneau,A., Primeau,M., Trudeau,L.E. & Michnick,S.W. Beta-lactamase protein fragment complementation assays as in vivo and in vitro sensors of protein protein interactions. *Nat. Biotechnol.* **20**, 619-622 (2002).

2. Remy,I., Ghaddar,G. & Michnick,S.W. Using the beta-lactamase protein-fragment complementation assay to probe dynamic protein-protein interactions. *Nat. Protoc.* **2**, 2302-2306 (2007).

3. Luker,K.E. *et al.* Kinetics of regulated protein-protein interactions revealed with firefly luciferase complementation imaging in cells and living animals. *Proc. Natl. Acad. Sci. U. S. A* **101**, 12288-12293 (2004).

4. Stefan,E. *et al.* Quantification of dynamic protein complexes using Renilla luciferase fragment complementation applied to protein kinase A activities in vivo. *Proc. Natl. Acad. Sci. U. S. A* **104**, 16916-16921 (2007).

5. Pelletier,J.N., Campbell-Valois,F.X. & Michnick,S.W. Oligomerization domain-directed reassembly of active dihydrofolate reductase from rationally designed fragments. *Proc. Natl. Acad. Sci. U. S. A* **95**, 12141-12146 (1998).

6. Remy,I., Campbell-Valois,F.X. & Michnick,S.W. Detection of protein-protein interactions using a simple survival protein-fragment complementation assay based on the enzyme dihydrofolate reductase. *Nat. Protoc.* **2**, 2120-2125 (2007).

7. Wehrman,T., Kleaveland,B., Her,J.H., Balint,R.F. & Blau,H.M. Protein-protein interactions monitored in mammalian cells via complementation of beta - lactamase enzyme fragments. *Proc. Natl. Acad. Sci. U. S. A* **99**, 3469-3474 (2002).

8. Hu,C.D. & Kerppola,T.K. Simultaneous visualization of multiple protein interactions in living cells using multicolor fluorescence complementation analysis. *Nat. Biotechnol.* **21**, 539-545 (2003).

9. Famulok,M., Hartig,J.S. & Mayer,G. Functional aptamers and aptazymes in biotechnology, diagnostics, and therapy. *Chem. Rev.* **107**, 3715-3743 (2007).

10. Hermann,T. & Patel,D.J. Adaptive recognition by nucleic acid aptamers. *Science* **287**, 820-825 (2000).

11. Hartig,J.S. *et al.* Protein-dependent ribozymes report molecular interactions in real time. *Nat. Biotechnol.* **20**, 717-722 (2002).

12. Serganov,A. & Patel,D.J. Ribozymes, riboswitches and beyond: regulation of gene expression without proteins. *Nat. Rev. Genet.* **8**, 776-790 (2007).

13. Scott,W.G. Ribozymes. *Curr. Opin. Struct. Biol.* **17**, 280-286 (2007).

14. Winkler,W.C., Nahvi,A., Roth,A., Collins,J.A. & Breaker,R.R. Control of gene expression by a natural metabolite-responsive ribozyme. *Nature* **428**, 281-286 (2004).

15. Teixeira,A. *et al.* Autocatalytic RNA cleavage in the human beta-globin pre-mRNA promotes transcription termination. *Nature* **432**, 526-530 (2004).

16. Tarasow,T.M., Tarasow,S.L. & Eaton,B.E. RNA-catalysed carbon-carbon bond formation. *Nature* **389**, 54-57 (1997).

17. Curtis,E.A. & Bartel,D.P. New catalytic structures from an existing ribozyme. *Nat. Struct. Mol. Biol.* **12**, 994-1000 (2005).

18. Johnston,W.K., Unrau,P.J., Lawrence,M.S., Glasner,M.E. & Bartel,D.P. RNA-catalyzed RNA polymerization: accurate and general RNA-templated primer extension. *Science* **292**, 1319-1325 (2001).

19. Muller,U.F. & Bartel,D.P. Improved polymerase ribozyme efficiency on hydrophobic assemblies. *RNA.* **14**, 552-562 (2008).

20. Akashi,H., Matsumoto,S. & Taira,K. Gene discovery by ribozyme and siRNA libraries. *Nat. Rev. Mol. Cell Biol.* **6**, 413-422 (2005).

21. Sullenger,B.A. & Gilboa,E. Emerging clinical applications of RNA. *Nature* **418**, 252-258 (2002).

22. Uhlenbeck,O.C. A small catalytic oligoribonucleotide. *Nature* **328**, 596-600 (1987).

23. Hasegawa,S., Gowrishankar,G. & Rao,J. Detection of mRNA in mammalian cells with a split ribozyme reporter. *Chembiochem.* **7**, 925-928 (2006).

24. Sioud,M. Ribozyme- and siRNA-mediated mRNA degradation: a general introduction. *Methods Mol. Biol.* **252**, 1-8 (2004).

25. Amontov,S.V. & Taira,K. Hammerhead minizyme with high cleavage activity: a dimeric structure as the active conformation of minizymes. *J. Am. Chem. Soc.* **118**, 1624-1628 (1996).

26. Amontov,S., Nishikawa,S. & Taira,K. Dependence on Mg2+ ions of the activities of dimeric hammerhead minizymes. *FEBS Lett.* **386**, 99-102 (1996).

27. Kuwabara,T., Amontov,S.V., Warashina,M., Ohkawa,J. & Taira,K. Characterization of several kinds of dimer minizyme: simultaneous cleavage at two sites in HIV-1 tat mRNA by dimer minizymes. *Nucleic Acids Res.* **24**, 2302-2310 (1996).

28. Tanabe,T. *et al.* Maxizymes, novel allosterically controllable ribozymes, can be designed to cleave various substrates. *Biomacromolecules.* **1**, 108-117 (2000).

29. Kuwabara,T. *et al.* A novel allosterically trans-activated ribozyme, the maxizyme, with exceptional specificity in vitro and in vivo. *Mol. Cell* **2**, 617-627 (1998).

30. Kuwabara,T., Warashina,M., Nakayama,A., Ohkawa,J. & Taira,K. tRNAVal-heterodimeric maxizymes with high potential as geneinactivating agents: simultaneous cleavage at two sites in HIV-1 Tat mRNA in cultured cells. *Proc. Natl. Acad. Sci. U. S. A* **96**, 1886-1891 (1999).

31. Tanabe,T. *et al.* Oncogene inactivation in a mouse model. *Nature* **406**, 473-474 (2000).

32. Breaker,R.R. Engineered allosteric ribozymes as biosensor components. *Curr. Opin. Biotechnol.* **13**, 31-39 (2002).

33. Hasegawa,S., Jackson,W.C., Tsien,R.Y. & Rao,J. Imaging Tetrahymena ribozyme splicing activity in single live mammalian cells. *Proc. Natl. Acad. Sci. U. S. A* **100**, 14892-14896 (2003).

34. Hasegawa,S., Choi,J.W. & Rao,J. Single-cell detection of trans-splicing ribozyme in vivo activity. *J. Am. Chem. Soc.* **126**, 7158-7159 (2004).

35. Zlokarnik,G. *et al.* Quantitation of transcription and clonal selection of single living cells with beta-lactamase as reporter. *Science* **279**, 84-88 (1998).

36. Gao,W., Xing,B., Tsien,R.Y. & Rao,J. Novel fluorogenic substrates for imaging beta-lactamase gene expression. *J. Am. Chem. Soc.* **125**, 11146-11147 (2003).

37. Villalobos,V., Naik,S. & Piwnica-Worms,D. Current state of imaging protein-protein interactions in vivo with genetically encoded reporters. *Annu. Rev. Biomed. Eng* **9**, 321-349 (2007).

38. Lee,J.F., Hesselberth,J.R., Meyers,L.A. & Ellington,A.D. Aptamer database. *Nucleic Acids Res.* **32**, D95-100 (2004).

39. Lebruska,L.L. & Maher,L.J., III. Selection and characterization of an RNA decoy for transcription factor NF-kappa B. *Biochemistry* **38**, 3168-3174 (1999).

40. Cassiday,L.A. & Maher,L.J., III. Yeast genetic selections to optimize RNA decoys for transcription factor NF-kappa B. *Proc. Natl. Acad. Sci. U. S. A* **100**, 3930-3935 (2003).

41. Huang,D.B. *et al.* Crystal structure of NF-kappaB (p50)2 complexed to a high-affinity RNA aptamer. *Proc. Natl. Acad. Sci. U. S. A* **100**, 9268-9273 (2003).

42. Cassiday,L.A. & Maher,L.J., III. In vivo recognition of an RNA aptamer by its transcription factor target. *Biochemistry* **40**, 2433-2438 (2001).

43. Li,X. & Stark,G.R. NFkappaB-dependent signaling pathways. *Exp. Hematol.* **30**, 285-296 (2002).

44. Ghosh,S., May,M.J. & Kopp,E.B. NF-kappa B and Rel proteins: evolutionarily conserved mediators of immune responses. *Annu. Rev. Immunol.* **16**, 225-260 (1998).

45. Moorthy,A.K., Huang,D.B., Wang,V.Y., Vu,D. & Ghosh,G. X-ray structure of a NF-kappaB p50/RelB/DNA complex reveals assembly of multiple dimers on tandem kappaB sites. *J. Mol. Biol.* **373**, 723-734 (2007).

46. Ghosh,G., van Duyne,G., Ghosh,S. & Sigler,P.B. Structure of NF-kappa B p50 homodimer bound to a kappa B site. *Nature* **373**, 303-310 (1995).

47. Kuwabara,T., Warashina,M., Nakayama,A., Ohkawa,J. & Taira,K. tRNAVal-heterodimeric maxizymes with high potential as geneinactivating agents: simultaneous cleavage at two sites in HIV-1 Tat mRNA in cultured cells. *Proc. Natl. Acad. Sci. U. S. A* **96**, 1886-1891 (1999).

48. Kawasaki,H., Warashina,M., Kuwabara,T. & Taira,K. Helicase-attached novel hybrid ribozymes. *Methods Mol. Biol.* **252**, 237-243 (2004).

49. Shimayama,T., Nishikawa,S. & Taira,K. Generality of the NUX rule: kinetic analysis of the results of systematic mutations in the trinucleotide at the cleavage site of hammerhead ribozymes. *Biochemistry* **34**, 3649-3654 (1995).

50. Sano,M. & Taira,K. Ribozyme expression systems. *Methods Mol. Biol.* **252**, 195-207 (2004).

51. Kuwabara,T., Warashina,M., Nakayama,A., Ohkawa,J. & Taira,K. tRNAVal-heterodimeric maxizymes with high potential as gene inactivating agents: simultaneous cleavage at two sites in HIV-1 Tat mRNA in cultured cells. *Proc. Natl. Acad. Sci. U. S. A* **96**, 1886-1891 (1999).

52. Knuefermann,P. *et al.* Myotrophin/V-1, a protein up-regulated in the failing human heart and in postnatal cerebellum, converts NFkappa B p50-p65 heterodimers to p50-p50 and p65-p65 homodimers. *J. Biol. Chem.* **277**, 23888-23897 (2002).

53. Sengchanthalangsy,L.L. *et al.* Characterization of the dimer interface of transcription factor NFkappaB p50 homodimer. *J. Mol. Biol.* **289**, 1029-1040 (1999).

54. Sivasubramanian,N., Adhikary,G., Sil,P.C. & Sen,S. Cardiac myotrophin exhibits rel/NF-kappa B interacting activity in vitro. *J. Biol. Chem.* **271**, 2812-2816 (1996).

55. Gupta,S. & Sen,S. Myotrophin-kappaB DNA interaction in the initiation process of cardiac hypertrophy. *Biochim. Biophys. Acta* **1589**, 247-260 (2002).

56. Lin,C.W. & Ting,A.Y. Transglutaminase-catalyzed site-specific conjugation of small-molecule probes to proteins in vitro and on the surface of living cells. *J. Am. Chem. Soc.* **128**, 4542-4543 (2006).

57. Kensch,O. *et al.* HIV-1 reverse transcriptase-pseudoknot RNA aptamer interaction has a binding affinity in the low picomolar range coupled with high specificity. *J. Biol. Chem.* **275**, 18271-18278 (2000).

58. Martell,R.E., Nevins,J.R. & Sullenger,B.A. Optimizing aptamer activity for gene therapy applications using expression cassette SELEX. *Mol. Ther.* **6**, 30-34 (2002).

59. Sano,M. & Taira,K. Hammerhead ribozyme-based target discovery. *Methods Mol. Biol.* **360**, 143-153 (2007).

60. Khvorova,A., Lescoute,A., Westhof,E. & Jayasena,S.D. Sequence elements outside the hammerhead ribozyme catalytic core enable intracellular activity. *Nat. Struct. Biol.* **10**, 708-712 (2003).

61. Canny,M.D. *et al.* Fast cleavage kinetics of a natural hammerhead ribozyme. *J. Am. Chem. Soc.* **126**, 10848-10849 (2004).

62. Martick,M. & Scott,W.G. Tertiary contacts distant from the active site prime a ribozyme for catalysis. *Cell* **126**, 309-320 (2006).

63. Sengchanthalangsy,L.L. *et al.* Characterization of the dimer interface of transcription factor NFkappaB p50 homodimer. *J. Mol. Biol.* **289**, 1029-1040 (1999).

64. Koseki,S., Ohkawa,J., Yamamoto,R., Takebe,Y. & Taira,K. A simple assay system for examination of the inhibitory potential in vivo of decoy RNAs, ribozymes and other drugs by measuring the Tat-mediated transcription of a fusion gene composed of the long terminal repeat of HIV-1 and a gene for luciferase. *J. Control Release* **53**, 159-173 (1998).

Chapter 3: Proximity biotinylation for the detection of endogenous protein-protein interactions in fixed samples

The work discussed in this chapter is unpublished. Optimization of the reaction conditions for the preparation of antibody-biotin ligase conjugates was performed in collaboration with Laura Martínez-Hernández. Part of the work described in this chapter has also been presented in her Master Thesis.

Introduction

In the previous chapter, we described our efforts to develop a new protein-protein interaction (PPI) reporter that could detect endogenous proteins in the live-cell context. The proposed methodology did however not hold up to its promise because of the strict geometric and interaction-affinity constraints. In this chapter, we decided to continue our quest for a method that detects endogenous protein-protein interactions (PPIs), but within a less challenging context, that of a fixed sample. Although the method described herein has the limitation of lacking temporal resolution, it should nonetheless display superior spatial resolution to the method described in Chapter 1, where the diffusion of the fluorescent β-lactamase substrate resulted in lack of subcellular compartmentalization.

Our proposed approach is a modification of the classic immunofluorescence (IF) technique. In IF, fluorophore-conjugated antibodies against the target proteins are used to stain a fixed sample, and the overlap between the two colors reports on protein co-localization[1]. At the time when this project was being pursued, IF was the only technique capable of providing spatial information on endogenous PPIs. However, the spatial resolution of IF is limited to ~ 250 nm in conventional microscopy[2]. Therefore, IF can only report on protein co-localization but cannot demonstrate whether an interaction is truly occurring. At the other end of the spatial resolution spectrum, the combination of fluorescence resonance energy transfer (FRET) and IF results in the detection of proteins that are separated by distances of 1–10 nm (Ref. 3). Our proposed methodology was thus an attempt to bridge the 250 nm spatial resolution of classic IF with the 10 nm resolution of IF-FRET. Figure 3-1 depicts our detection scheme, where the two fluorophores used in IF have been replaced by an enzyme-substrate pair. Here, each antibody is conjugated to either the bacterial enzyme biotin ligase (BirA) or its 15-amino acid peptide substrate (Figure 3-1A). *E. coli* biotin ligase is a 35 kD enzyme that catalyzes the ATP-dependent covalent ligation of biotin to a lysine sidechain of a 15-amino acid recognition sequence called the "acceptor peptide" (AP)[4] (Figure 3-1B). Upon addition of biotin to cells, BirA ligates biotin to the AP only if A and B are in close proximity. Subsequently, biotinylated AP can be detected with extremely high sensitivity and specificity on fixed cells by streptavidin staining[5].

We chose BirA-AP as our enzyme-substrate pair because our laboratory has extensive knowledge of the system, and because Irwin Chen and Mark howarth, in our lab, had previously established the use of BirA-AP for site-specific protein labeling in the cellular context[6,7]. We and others have also shown that BirA can biotinylate the AP when it is placed in different contexts such as the N- and C-terminus of a protein or in an internal loop. Additionally, the sequence specificity of BirA is extremely high, as it has been demonstrated that it does not biotinylate any endogenous mammalian proteins[7,8]. Both properties were important considerations in our system, as they suggest that the recognition of AP by BirA is not context dependent, and that BirA addition to fixed samples would result in biotinylation of AP but no other protein, thereby eliminating a source of false positives.

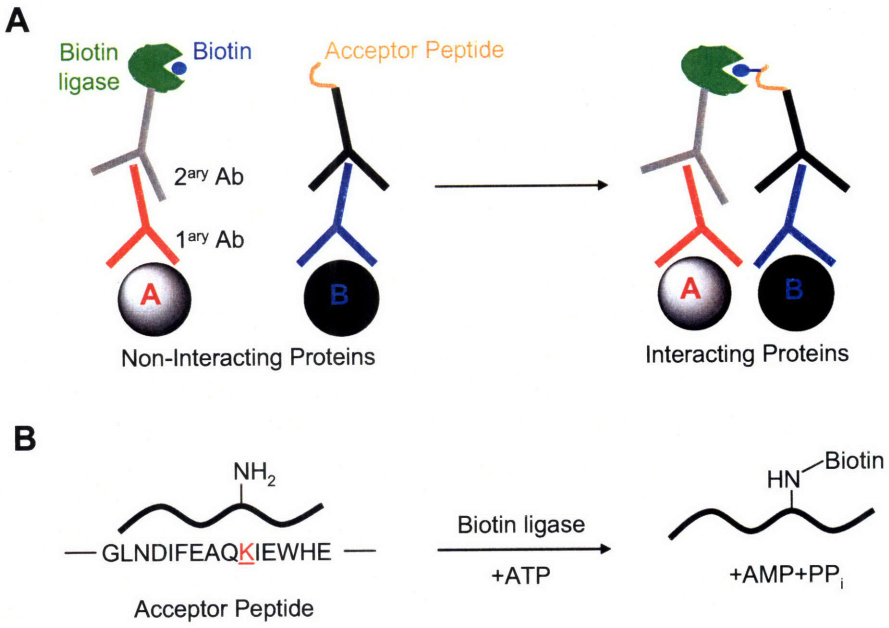


Figure 3-1: Proposed scheme for the detection of endogenous protein-protein interactions in fixed cells. (**A**) Proximity biotinylation for the immunodetection of endogenous protein-protein interactions. Each target protein, A and B, is recognized by a specific primary antibody (red or blue), which is in turn bound by a secondary antibody (grey or black) conjugated to either the *E. coli* enzyme biotin ligase (green) or its 15-amino acid peptide substrate sequence (orange). If A and B are interacting before cell fixation, then biotin ligase will be in close proximity to the acceptor peptide, and biotin addition will result in biotinylation of the peptide substrate. In contrast, if the target proteins are not interacting partners, biotin ligase will no be able to bind and biotinylate the acceptor peptide. (**B**) *E. coli* biotin ligase catalyzes the ATP-dependent ligation of biotin to the ε-amino group of a lysine sidechain within a 15-amino acid peptide sequence, called the acceptor peptide (AP).

Results

We have divided the work performed towards the development of this methodology into three main sections: first, development of robust protocols to chemically conjugate biotin ligase and the acceptor peptide to immunoglobulins; second, testing of the binding capability and enzymatic activity of the antibody conjugates; and third, testing the antibody conjugates for immunodetection of endogenous protein-protein interactions in live cells and in fixed biological samples.

Strategies for antibody derivatization

The unique structural characteristics of antibody molecules offer a number of choices for modification and conjugation schemes. A detailed illustration of the structure of an antibody is shown in Figure 3-2. The most basic immunoglobulin (Ig) molecule is composed of two identical light chains (~ 25 kD each chain) and two identical heavy chains (~ 50–75 kD each chain), which are held together by non-covalent interactions as well as a number of disulfide bonds[9]. The light chains are disulfide-bonded to the heavy chains in the C_L and C_{H1} regions, whereas the heavy chains are disulfide-bonded to each other in the hinge region. There are two antigen binding sites on each IgG, which are formed by the proximity of the heavy and light chains at the N-terminus of the molecule, called the hypervariable region[9]. The antigen binding site is determined not by the linear sequence of amino acids on each chain, but mostly by the unique orientation of these groups in the three-dimensional space.

Figure 3-2 highlights the antibody functional groups that are suitable for conjugation. First, as with any protein, N-terminal and lysine primary amines can be targeted for crosslinking by amine-reactive groups such as activated esters, ketones, and aldehydes. One of the most commonly used amine-reactive reagents is *N*-hydroxysuccinimide (NHS). Conversely, carboxylate groups present at the C-terminus of the antibody chains and the sidechains of glutamic and aspartic acid residues can be activated for example with 1-ethyl-3-[3-dimethylaminopropyl]carbodiimide hydrochloride (EDC) and then coupled to amine, hydrazide, or hydroxylamine functional groups. Third, the carbohydrate molecules present in most (but not all) antibodies can be

oxidized to generate aldehydes, which can then react with exogenous amines. Finally, the disulfides that hold the heavy chains together at the hinge region can be selectively cleaved with a reducing agent to create two half-antibody molecules, each containing an antigen binding site. This preparation exposes free thiols that can be targeted for conjugation using sulfhydryl-reactive probes such as maleimides or haloacetamides. Out of the four conjugation schemes, the last two are usually preferred because they employ functional groups (carbohydrates and sulfhydryls) present only at specific positions of the immunoglobulin surface and away from the antigen binding site, and are therefore more likely to preserve antibody binding capability[10]. Thus, with the appropriate selection of the conjugation chemical reaction and knowledge of antibody structure, the immunoglobulin molecule can be modified so that its antigen binding potential remains available after conjugation to biotin ligase or the acceptor peptide.

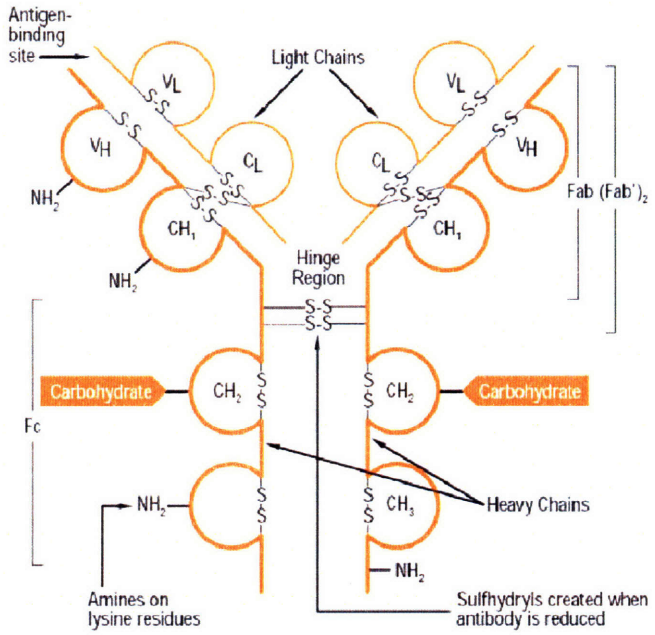


Figure 3-2: Detailed structure of an immunoglobulin G antibody molecule. Each antibody molecule is composed of two light chains (L subscripts) and two heavy chains (H subscripts). Each chain has a variable region (V_L and V_H) and one or more constant regions (C_L and C_{H1}, C_{H2}, etc.). Each antigen binding site is formed by one heavy chain and one light chain (the Fab – fragment antigen binding – arms). Fc (fraction crystallisable) represents the most conserved part of the antibody and it is not involved in antigen binding. Functional groups targeted by the different bioconjugation strategies (e.g., lysine or cysteine residues) are highlighted. Image taken from the Pierce website.

Chemical conjugation of the acceptor peptide to IgGs

We conjugated the acceptor peptide to anti-rabbit and anti-mouse immunoglobulins by employing their carbohydrates (Figure 3-3A). Antibody sugars were first oxidized using sodium periodate and the resulting aldehydes were immediately reacted with an excess of acceptor peptide containing a C-terminal hydrazide functionality. The formed hydrazone was then reduced to a hydrazide using sodium cyanoborohydride.

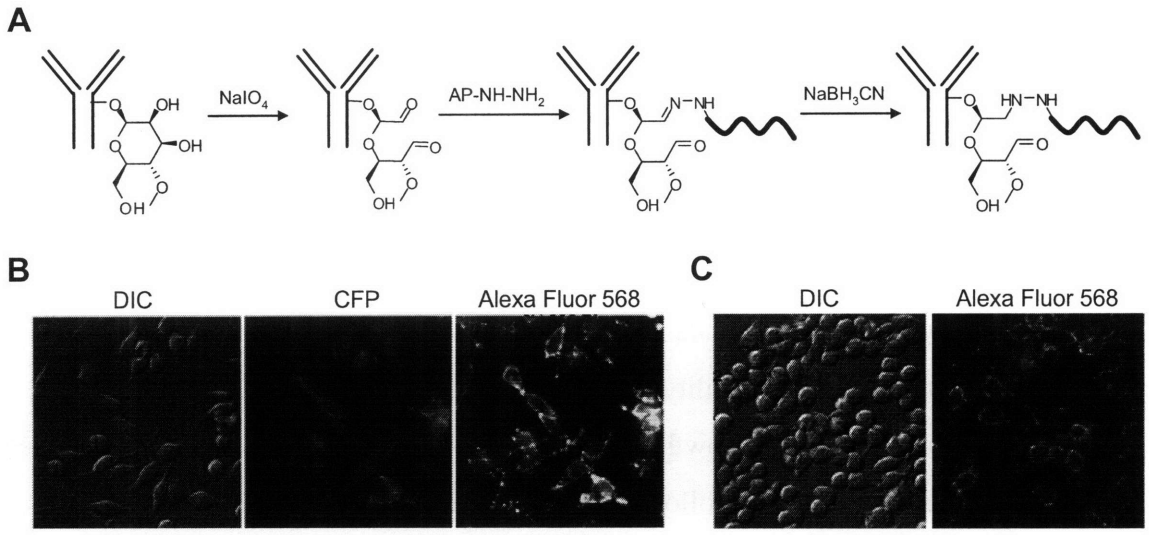


Figure 3-3: Conjugation of the acceptor peptide to immunoglobulins. (**A**) Conjugation scheme. Antibody sugars were oxidized with sodium periodate and the resulting aldehydes were reacted with a hydrazide-derivatized acceptor peptide. The formed hydrazone was reduced to a hydrazide using sodium cyanoborohydride. (**B**) Cell-surface detection of anti-rabbit-AP conjugates in live HeLa cells. An HA-CFP fusion was targeted to the cell surface using a transmembrane (TM) domain. Cell-surface HA was detected using rabbit anti-HA, followed by the anti-rabbit-acceptor peptide conjugate. The AP was then biotinylated by addition of soluble BirA, biotin, and ATP, and the introduced biotin was then detected with streptavidin conjugated to Alexa Fluor 568. Live cell images of the introduced Alexa Fluor 568 are shown to the right of the CFP and DIC images, which highlight the transfected cells. (**C**) Cell-surface detection of anti-mouse-AP conjugates in live HeLa cells. Endogenous epidermal growth factor receptor (EGFR) was detected using mouse anti-EGFR, followed by the anti-mouse-acceptor peptide conjugate. The AP was then biotinylated and detected as in B. Live cell images of the introduced Alexa Fluor 568 are shown to the right of the DIC image.

Optimization of the oxidation reaction was performed using fluorescein-hydrazide as a model small-molecule, as it can be more easily detected than the acceptor peptide. Although antibody oxidation is typically performed at acidic pH[10] to favor the subsequent reaction with the hydrazide, we found that pH 7.2 resulted in a significant degree of sugar

oxidation while leaving the antibody intact, whereas pH 5.5 resulted in ~ 50% loss of antigen binding capability (data not shown). Using fluorescein-hydrazide, we also determined that the oxidation reaction was done after 30 minutes in the presence of 1 mM sodium periodate (i.e., longer incubation times did not result in an increase of the fluorescein signal). Excess sodium periodate was then removed by gel filtration using Sephadex™ G25 resin.

In the second step of the conjugation, a hydrazide-derivatized synthetic acceptor peptide was added in excess (typically 1000-fold) of antibody and reacted for ~ 16 hours at pH 5. We tested the conjugation extent by *in vitro* biotinylation of AP followed by Western blot detection using streptavidin-HRP (data not shown). Although we had expected excess AP to be removed from cells during the washes, we later found that the AP-hydrazide was very sticky in fixed cells (data not shown), probably owing to its reactivity with the paraformaldehyde fixative. Thus, we thereon purified every batch of antibody-acceptor peptide conjugate by gel filtration.

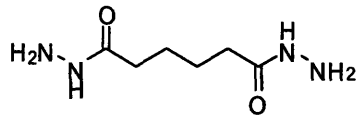
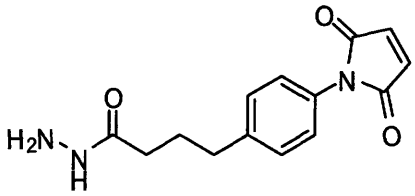
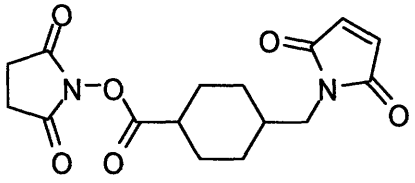
Every new batch of conjugates was also tested on a live-cell assay to confirm retention of antigen binding capability. We stained either a cell-surface HA epitope (Figure 3-3B) or the epidermal growth factor receptor (Figure 3-3C) with rabbit or mouse primary antibodies, respectively, followed by the anti-rabbit- or anti-mouse-acceptor peptide conjugates. After washing excess antibody conjugates, the bound molecules were biotinylated using soluble biotin ligase, and ligated biotin was detected using streptavidin conjugated to Alexa Fluor 568. Figure 3-3B and C shows that the antibody conjugates retained their binding capability as demonstrated by the streptavidin-Alexa Fluor 568 signal. Specificity controls were performed with the primary antibody omitted, and showed no labeling (data not shown). The lower signal observed for the anti-mouse conjugates (Figure 3-3C) most probably reflects the lower expression of endogenous EGFR compared to that of the over-expressed HA cell-surface epitope.

Chemical conjugation of biotin ligase to IgGs

Conjugation of biotin ligase to immunoglobulins turned out to be more difficult than that of the acceptor peptide. We hypothesized that the main obstacle was a low

conjugation yield caused by low BirA concentrations in the reaction. For example, whereas we routinely added a 1000-fold excess of acceptor peptide over antibody (see above), we could never achieve an excess larger than 20-fold for BirA. Additionally, BirA derivatization also proved challenging as enzymatic activity was lost in many cases. Table 3-1 summarizes the various strategies that we tested in order to maximize the conjugation yield, while preserving BirA enzymatic activity and antibody binding capability.

Table 3-1: Summary of strategies for conjugation of biotin ligase to immunoglobulins. MPBH: 4-(4-*N*-maleimidophenyl) butyric acid hydrazide. SMCC: succinimidyl 4-[*N*-maleimidomethyl]cyclohexane-1-carboxylate..

	Reactive Group 1		**Reactive Group 2**		**Chemical structure**
	Name	**Reacts towards**	**Name**	**Reacts towards**	
Reductive Amination	Amine	Activated esters or carbonyls	NA	NA	NA
Adipic dihydrazide	Hydrazide	Activated esters or carbonyls	Hydrazide	Activated esters or carbonyls	
MPBH	Hydrazide	Activated esters or carbonyls	Maleimide	Thiols	
SMCC	*N*-hydroxy succinimide	Amines	Maleimide	Thiols	

We first tried direct conjugation of BirA to the antibody via reductive amination, but this method mostly resulted in antibody self-crosslinking. We then tested the crosslinking reagent adipic dihydrazide, but its homobifunctional nature translated into a

high degree of BirA oligomerization. Finally, we tested two different heterobifunctional crosslinkers, among which only SMCC (succinimidyl 4-[*N*-maleimidomethyl]cyclohexane-1-carboxylate) yielded antibody-biotin ligase conjugates of controllable stoichiometry and in sufficient quantity. The following sections describe in detail the results obtained for each strategy.

Reductive amination results in antibody self-crosslinking

The simplest conjugation scheme involved reaction at high pH of the primary amines in biotin ligase (e.g., N-terminal amine and lysine ε–amines) with the aldehydes generated by oxidation of the antibody sugars (Figure 3-4A). The reaction product, an imine, was stabilized by reduction to a secondary amine with sodium cyanoborohydride. Figure 3-4B shows the results obtained for a typical conjugation reaction performed by reductive amination. We initially thought that the reaction was yielding a fair amount of antibody-biotin ligase conjugates (green arrows), as it could be inferred by comparing lanes 1 (BirA only) and 4 (antibody only) with lane 3 (conjugation reaction) in the Silver staining image. However, we later found that the new protein bands observed were the result of antibody self-crosslinking, and that their appearance did not require the presence of biotin ligase (compare lanes 2 and 3 in the Silver staining image and α-goat IgG blot).

Although we had anticipated that the antibody molecules could self-conjugate through the reaction of the sugar aldehydes with antibody's own amines during the high-pH reductive amination, we had expected that an excess of biotin ligase could reduce self-crosslinking to a minimum. Unfortunately, we found that the maximum achievable concentration of biotin ligase in the conjugation reaction (~ 20 μM) was not enough to prevent antibody self-crosslinking. We thus hypothesized that conjugation with periodate-oxidized antibodies would work best if the receiving biotin ligase molecules were previously modified to contain a hydrazide, and conjugation was done at moderate pH. It is worth noting that we also observed a certain degree of biotin ligase self-crosslinking (highlighted by red boxes in the α-His_6 and α-goat blots of Figure 3-4B), which could not be readily explained.

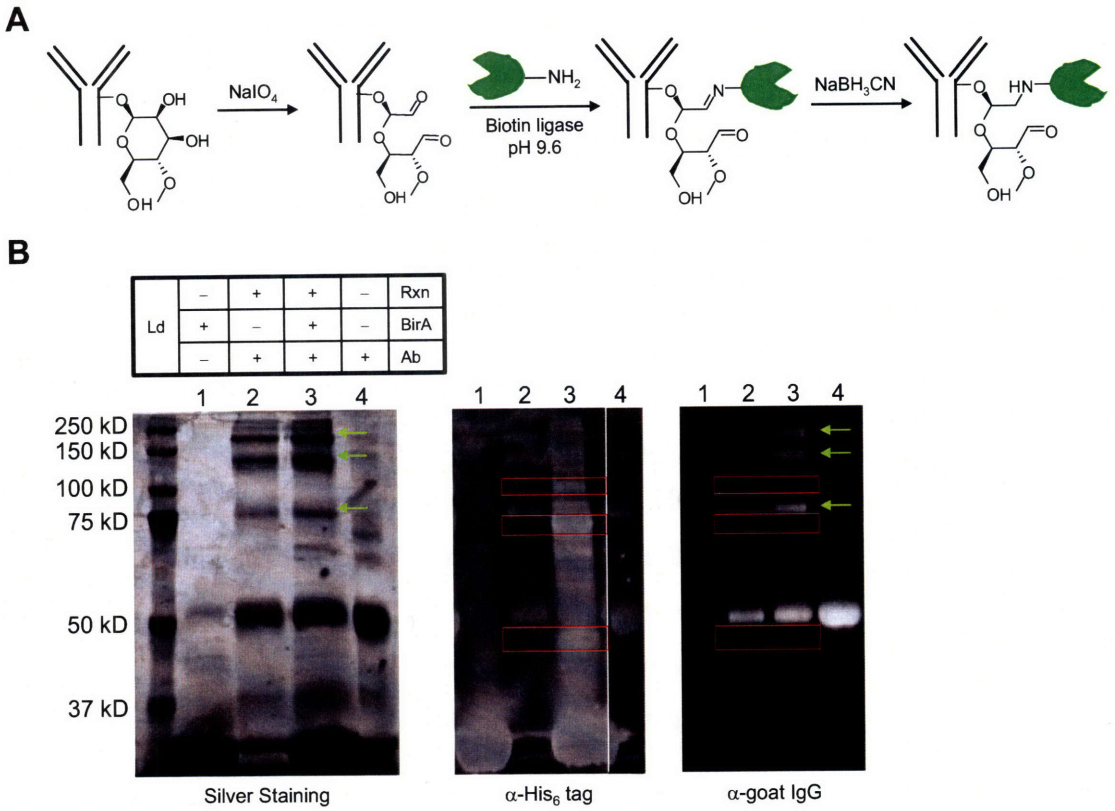


Figure 3-4: Conjugation of biotin ligase to immunoglobulins by reductive amination. (**A**) Conjugation scheme. Antibody sugars were oxidized with sodium periodate and the resulting aldehydes reacted with the primary amines present in biotin ligase. Conjugation reaction was performed at high pH to favor nucleophilic attack by the amines. The formed imine was reduced to an amine using sodium cyanoborohydride. (**B**) Analysis of the conjugation products by Western blotting. Oxidized antibody (~ 1 μM) was reacted with 10 μM BirA in sodium carbonate buffer, pH 9.6, followed by reduction with sodium cyanoborohydride. The reaction products were run on a 10% SDS-PAGE, and the extent of antibody-biotin ligase conjugation was analyzed by Silver staining and by α-His_6 blotting (to detect only biotin ligase) and α-Goat IgG blotting (to detect only the antibody). The conjugation reaction (lane 3) is compared side-by-side with a negative control where no biotin ligase was added during the reductive amination (lane 2), and with the biotin ligase and antibody reagents before conjugation (lanes 1 and 4, respectively). Red boxes highlight conjugation products containing biotin ligase but no antibody. Green arrows point to the products of antibody self-crosslinking.

Use of adipic-dihydrazide results in variable conjugate stoichiometry

We chose adipic dihydrazide (Table 3-1) as the crosslinking reagent to introduce a hydrazide functionality in biotin ligase. Figure 3-5A depicts the conjugation scheme. The carboxylate groups in biotin ligase were activated using 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride (EDC) and reacted with adipic

dihydrazide (ADH). The hydrazide-containing BirA was then combined with oxidized antibody and the reaction was allowed to proceed at neutral pH to form the hydrazone, which was then reduced to a hydrazide using sodium cyanoborohydride. During the conjugation reaction, BirA-hydrazide was always present in excess of antibody to ensure that each antibody was conjugated to none, one, or more BirA molecules, but that no two antibody molecules were linked together. This is an important consideration because antibody crosslinking will ultimately result in large polymers, which could expand long distances, thereby not reporting on truly interacting proteins.

Initial attempts to perform conjugation following this scheme failed due to both, incomplete removal of excess ADH, and low protein concentrations. The first problem was easily solved by substituting the dialysis after the EDC/adipic dihydrazide reaction with either Ni-NTA purification or gel filtration using a Sephadex™ G25 resin. The latter was usually preferred owing to its simplicity and because its shorter protocol contributed to prevent biotin ligase self-conjugation. We then studied the effect of protein concentration on the yield of the conjugation reaction. We found, for example, that whereas biotin-cadaverine (5 mM) could be efficiently conjugated to a 130 μM solution of oxidized invertase (a highly glycosylated 60 kD protein), no conjugation could be detected under the same conditions to 13 μM of antibody solution (maximum concentration achievable) (data not shown). Furthermore, using biotin-hydroxylamine, a probe that had been previously proven highly reactive towards oxidized antibodies[11], we found that we could only observe antibody biotinylation when the biotin-hydroxylamine was added at concentrations higher than 100 μM (data not shown). This finding explains the success of the conjugation of acceptor peptide to antibody and the failure of the conjugation to biotin ligase, for which we could never achieve concentrations higher than 20–30 μM.

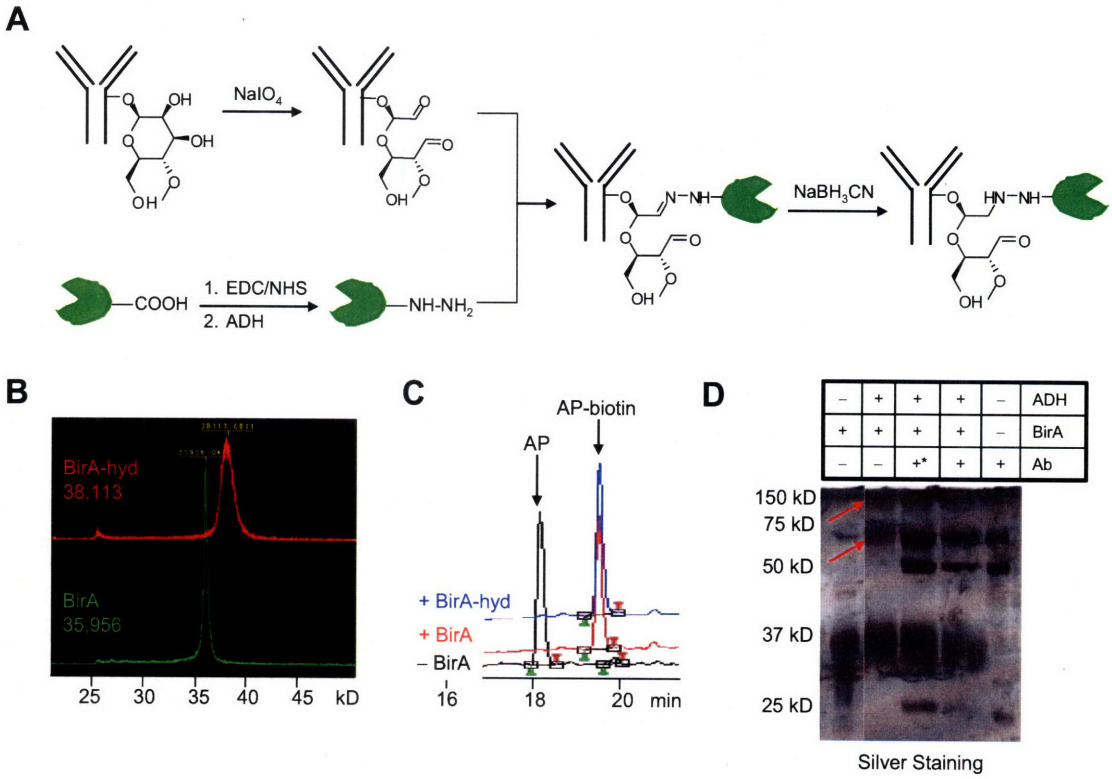


Figure 3-5: Conjugation of biotin ligase to immunoglobulins using adipic dihydrazide. (**A**) Conjugation scheme. First, antibody sugars were oxidized with sodium periodate to generate free aldehydes. In parallel, carboxylate groups in biotin ligase were activated with 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride (EDC) and reacted with adipic dihydrazide (ADH) to generate BirA-hydrazide. BirA-hydrazide was then combined with the oxidized antibody to generate antibody-BirA conjugates covalently bound *via* a hydrazone linkage. Finally, the formed hydrazone was reduced to a hydrazide using sodium cyanoborohydride. (**B**) Analsysis of the extent of hydrazide modification of biotin ligase by MALDI-TOF. 40 μM biotin ligase was EDC-activated and then combined with 40 mM ADH (red) or nothing (green) at room temperature overnight. The reaction mixtures were mixed with matrix in a 1:10 ratio and loaded directly onto the MALDI target. (**C**) HPLC activity assay showing that BirA-hydrazide retains catalytic activity. BirA-hydrazide was synthesized as in B, and then combined with 100 μM AP, 1 mM biotin, and 4 mM ATP for 20 minutes at room temperature. Negative controls with no BirA or un-modified BirA are also shown. (**D**) Analysis of the conjugation products by SDS-PAGE. Oxidized antibody (~ 1 μM) was reacted with 10 μM BirA-hydrazide in MES buffer, pH 6.2, followed by reduction with sodium cyanoborohydride. The reaction products were run on a 10% SDS-PAGE, and the extent of antibody-biotin ligase conjugation was analyzed by Silver staining. The conjugation reaction (lane 4) is compared side-by-side with negative controls with biotin ligase omitted (lane 2) or with un-oxidized antibody (lane 3). The biotin ligase and antibody reagents before conjugation are also shown for comparison (lanes 1 and 5, respectively). Red arrows point to BirA oligomers.

In parallel, we also tried to maximize the yield of hydrazide incorporation on biotin ligase. We found that incubation of biotin ligase with 50 mM adipic dihydrazide for 4 hours at room temperature resulted in modification of 100% of the BirA molecules,

as assessed by mass-spectrometry (Figure 3-5B). Importantly, the extent of the modification under these conditions was sufficiently high that reaction with the small-molecule fluorescein-monoaldehyde resulted in detectable biotin ligase-fluorescence signal (data not shown). We also confirmed, using an HPLC activity assay[6], that this large extent of modification did not affect BirA specific activity (Figure 3-5C). Lower concentrations of ADH resulted in a significant percentage of unmodified BirA molecules, whereas higher ADH concentrations or longer reaction times yielded catalytically inactive BirA, probably caused by modification of glutamate and/or aspartic residues near the active site[12].

After optimizing the synthesis of BirA-hydrazide conjugates, we attempted to perform the conjugation reaction. Figure 3-5D shows that, although higher molecular weight bands are observed in the conjugation reaction (lane 3), the same bands are present in the negative control with antibody omitted. This finding confirmed that the homobifunctional nature of the adipic dihydrazide molecule resulted in the formation of BirA oligomers. Variations in the BirA:Ab ratio or the BirA-hydrazide purification method (i.e., to minimize the time in between removal of excess ADH and addition of antibody) avoided BirA oligomerization to some extent, but did not result in the formation of a detectable amount of BirA-antibody conjugates. It must be noted that BirA has a natural dimerization affinity[13], therefore increasing even further the effective concentration for BirA self-conjugation.

SMCC-mediated conjugation offers highest yield and most controllable stoichiometry

In order to avoid biotin ligase or antibody self-conjugation, we invoked the use of heterobifunctional crosslinkers.

We first tested 4-(4-*N*-maleimidophenyl) butyric acid hydrazide (MPBH) because we had previously shown that the hydrazide functional group present in adipic dihydrazide reacts very efficiently with EDC activated BirA (see previous section). MPBH is a heterobifunctional crosslinker containing maleimide and hydrazide moieties (Table 3-1). Maleimides react with free sulfhydryls at pH 6.5–7.5 to form stable thioether bonds, whereas hydrazide groups react with activated esters or carbonyls to form stable hydrazone bonds. As previously mentioned, it was favorable to keep BirA always in

excess of antibody during the conjugation reactions to avoid antibody crosslinking. For this reason, we first activated the carboxylate groups in BirA with EDC and then reacted them with MPBH to generate BirA-maleimide conjugates. Unfortunately, for reasons that are not yet clear, the extent of BirA modification by the MPBH hydrazide moiety was never higher than ~ 15%, much lower than that obtained with ADH. We also tested the conjugation of MPBH to the sugar aldehydes obtained by antibody oxidation, as it had been previously reported[14], but the modification extent was also very low. The low reactivity of the hydrazide moiety in MPBH thus limited the usefulness of this reagent.

We next switched our attention to what is probably the most popular of all the heterobifunctional crosslinkers: SMCC. Succinimidyl-4-(*N*-maleimidomethyl) cyclohexane-1-carboxylate (SMCC) is a non-cleavable crosslinker that contains an amine-reactive *N*-hydroxysuccinimide ester (NHS ester) and a sulfhydryl-reactive maleimide group. NHS is an acylating agent that can undergo nucleophilic substitution with primary amines to form stable amide bonds at pH 7–9. The active NHS ester chemistry is characteristically the most labile functional group of SMCC, being susceptible to rapid hydrolysis under the aqueous conditions of a bio-conjugation reaction. The sulfhydryl-reactive group, however, is usually much more stable to breakdown in aqueous environments. Furthermore, the maleimide group of SMCC is unusually stable up to pH 7.5 because of the cyclohexane bridge in the spacer arm. Therefore, SMCC is usually employed in multi-step conjugations where the amines of one protein first react with the NHS ester and the resulting protein-maleimide is then reacted with a sulfhydryl containing biomolecule[15,16].

Our initial conjugation scheme is depicted in Figure 3-6A. The disulfides bonds in the hinge region of the antibody were reduced with dithiothreitol (DTT) to liberate the free thiols, which were then reacted with maleimide-conjugated biotin ligase to afford the biotin ligase-antibody conjugates. Specifically, activation of 10 μM BirA with 1 mM SMCC was performed in 0.1 M sodium carbonate buffer at pH 8.2 for 30 minutes at 37 °C. Reduction of the antibody was done at 6.7 μM antibody and 20 mM DTT. Excess un-reacted SMCC and DTT was removed by gel filtration using a Sephadex™ G25 resin, and the purified BirA-SMCC and reduced antibody were mixed in a 5:1 ratio and allowed to react for 1.5 hours at room temperature. The reaction products were then analyzed by

SDS-PAGE and Western blotting as described in the experimental section. A typical result is shown in Figure 3-6B. We found that SMCC resulted in a larger extent of conjugation than any of the previous strategies. Comparison of lanes 3, 4, and 5 confirms the appearance of higher-molecular weight bands (highlighted with a red box), which contained both biotin ligase and antibody. However, three main problems were encountered with this conjugation scheme.

First, when we tested the conjugates in an HPLC activity assay, they proved unable to perform biotinylation (data not shown). We hypothesized that the loss of BirA activity was happening during the initial reaction with SMCC, which could be acylating BirA's catalytic lysine. Lysine 183 is believed to be involved in the formation of the biotin-AMP ester, and is thus essential for BirA catalytic activity[12]. We tested our hypothesis by pre-incubating BirA with biotin-AMP before addition of SMCC. It has been reported that BirA displays extremely high affinity for the adenylate ester of biotin ($k_{off} \sim 2.7 \times 10^{-4}$ s^{-1}, $K_d \sim 50$ pM[17]) and it is thus expected that, once bound, biotin-AMP is not released from the active site unless transferred to the protein substrate. Figure 3-6C compares the biotinylation activity of un-modified BirA (black trace) with that of BirA activated with SMCC in the presence or absence of biotin-AMP (purple and blue traces, respectively). As expected, binding of the biotin-AMP ester partially protected the active site of BirA. However, complete protection of BirA's active site required milder reaction conditions. To find optimal conditions for the synthesis of BirA-maleimide conjugates, we used a combination of HPLC assays, to assess BirA activity, and mass-spectrometry (MALDI-TOF), to estimate the extent of BirA modification. MALDI-TOF analysis confirmed that incubation with just 200 μM SMCC instead of 1 mM resulted in modification of 100% of the BirA molecules (35,125.16 D for un-modified BirA *vs.* 36,072.72 D for SMCC-BirA). Under these conditions, catalytic activity was completely preserved in the BirA-SMCC conjugates if pre-incubation with biotin-AMP was performed.

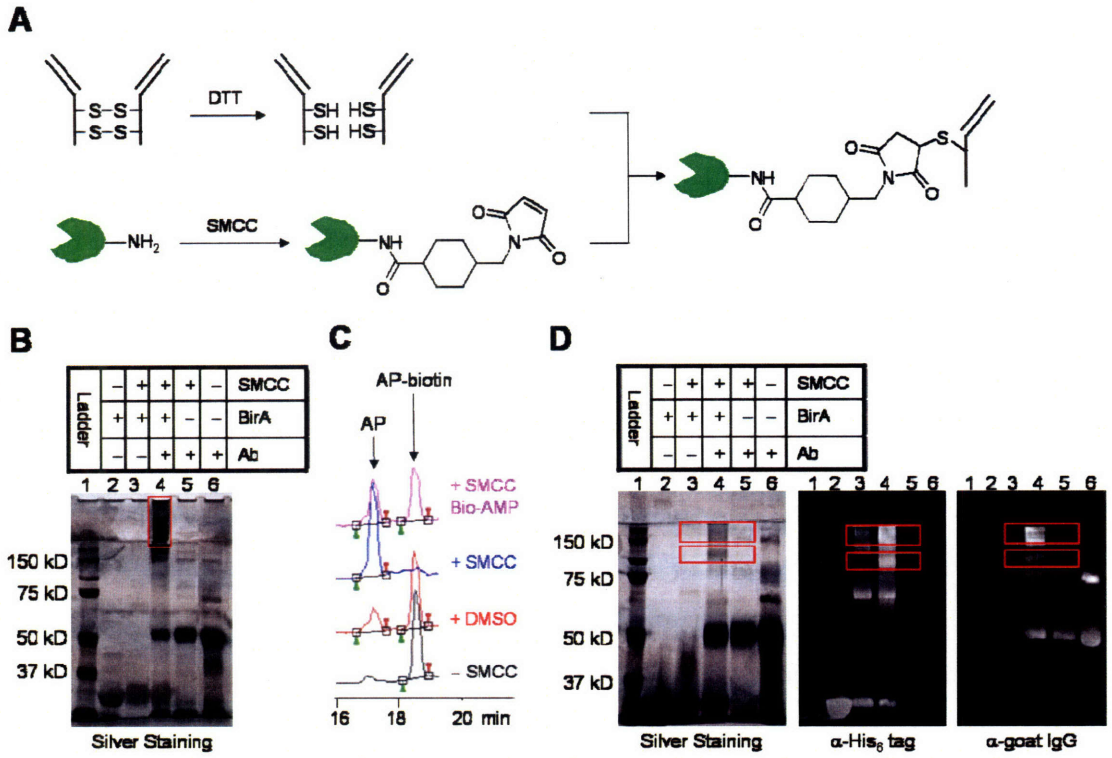


Figure 3-6: Conjugation of biotin ligase to immunoglobulins via BirA acylation by SMCC. (**A**) Conjugation scheme. First, antibody disulfide bonds were reduced with dithiothreitol (DTT) to generate free thiols in the hinge region. In parallel, lysine residues in biotin ligase were reacted with the NHS-ester of SMCC to generate BirA-maleimide conjugates. Reduced antibody was then combined with BirA-maleimide to generate antibody-BirA conjugates. (**B**) Analysis of the conjugation products by SDS-PAGE. BirA (~ 10 μM) was reacted with 1 mM SMCC in 0.1 M sodium carbonate buffer, pH 8.2 for 30 minutes at 37 ºC. Reduction of the antibody was done at 6.7 μM antibody and 20 mM DTT. Reduced antibody was combined with biotin ligase in a 1:5 ratio and incubated for 1.5 hours at room temperature. The reaction products were run on a 10% SDS-PAGE, and the extent of antibody-biotin ligase conjugation was analyzed by Silver staining. The conjugation reaction (lane 4) is compared side-by-side with negative controls where no biotin ligase or no antibody was added (lanes 3 and 5, respectively). The biotin ligase and antibody reagents before conjugation are also shown for comparison (lanes 2 and 4, respectively). Red box highlights BirA-antibody conjugates. (**C**) HPLC activity assay showing that BirA-maleimide looses catalytic activity. BirA-maleimide was synthesized as in B, and then combined with 100 μM AP, 1 mM biotin, and 4 mM ATP for 20 minutes at room temperature (blue trace). Negative controls with un-modified BirA or just DMSO are also shown (black and red traces, respectively). Pre-incubation of BirA with biotin-AMP partially protects the active site yielding a % of the BirA population that is catalytically active (purple trace). (**D**) Same as B but BirA-maleimide was synthesized under milder reactions conditions. Specifically, ~ 10 μM BirA was pre-incubated with 10 μM biotin-AMP for 5 minutes at room temperature before addition of 1 mM *N*-ethylmaleimide (NEM). After 1 hour, SMCC was added to C_f = 200 μM followed by incubation in sodium phosphate buffer, pH 7.1 for 1.5 hours at room temperature. Red boxes highlight BirA-antibody conjugates.

The second problem we encountered was the fact BirA-antibody conjugates predominantly formed higher order agglomerates (Figure 3-6B). It is worth noting that the conjugates remained in the stacking part of the gel, at molecular weights > 200 kD, as highlighted by the red box in Figure 3-6B. We hypothesized that the high molecular weight observed was caused by BirA oligomerization during the SMCC reaction. Oligomerization of BirA could have two distinct origins. First, at a reaction pH of 8.2, the maleimide moiety could be reacting with BirA lysine residues. We had initially chosen pH 8.2 because the NHS-ester acylation reaction proceeds faster at higher pH. However, although at pH 7 the reaction of the maleimide with sulfhydryls proceeds at a rate 1000 times greater than its reaction with amines, at more alkaline pH values, its reaction with amines becomes more evident. Second, regardless of the pH, the maleimide moiety in SMCC could be reacting with the unique cysteine in BirA. We tested this hypothesis by capping BirA only cysteine (Cys107) with *N*-ethylmaleimide (NEM). We performed the SMCC reaction at a range of pH 7.0–8.2 both in the presence and absence of NEM, and analyzed the reaction products by Western blotting detection of the His_6 tag in BirA. We found that BirA oligomerization could be avoided if NEM was added in a 10-fold excess and previous to SMCC addition, but only when the SMCC reaction was performed at pH < 7.5 (data not shown). These results confirmed that, at higher pH, the maleimide moiety was indeed reacting with the lysine residues.

Figure 3-6D illustrates the outcome of incorporating the previous modifications into the conjugation procedure. Comparison of Figure 3-6B and D confirmed that lower pH, combined with NEM capping of BirA Cys107 and milder SMCC reactions conditions, resulted in antibody-BirA conjugates of a more defined stoichiometry, corresponding to the covalent linkage of one antibody heavy chain with 1–3 BirA molecules (90–200 kD). HPLC activity tests confirmed that the conjugates retained catalytic activity.

The third problem encountered using this conjugation could however not be solved. To our surprise, when we tested the antibody-BirA conjugates in cells, we found that they were self-biotinylated. Biotin signal was observed when an anti-rabbit-BirA conjugate was added to anti-HA stained cells, even in the absence of any AP and/or exogenous biotin (data not shown). Subsequent tests by WB confirmed that self-

biotinylation was happening during the BirA activation with SMCC. We attempted to avoid self-biotinylation by reducing the concentration of biotin-AMP that we added during the SMCC reaction to protect the catalytic lysine, followed by extensive dialysis to remove excess biotin-AMP. Unfortunately, none of these completely removed the biotin signal. We then decided to reverse the order of the reaction steps, such that we combined BirA with the maleimide moiety in SMCC, thus eliminating the need for protecting the catalytic lysine.

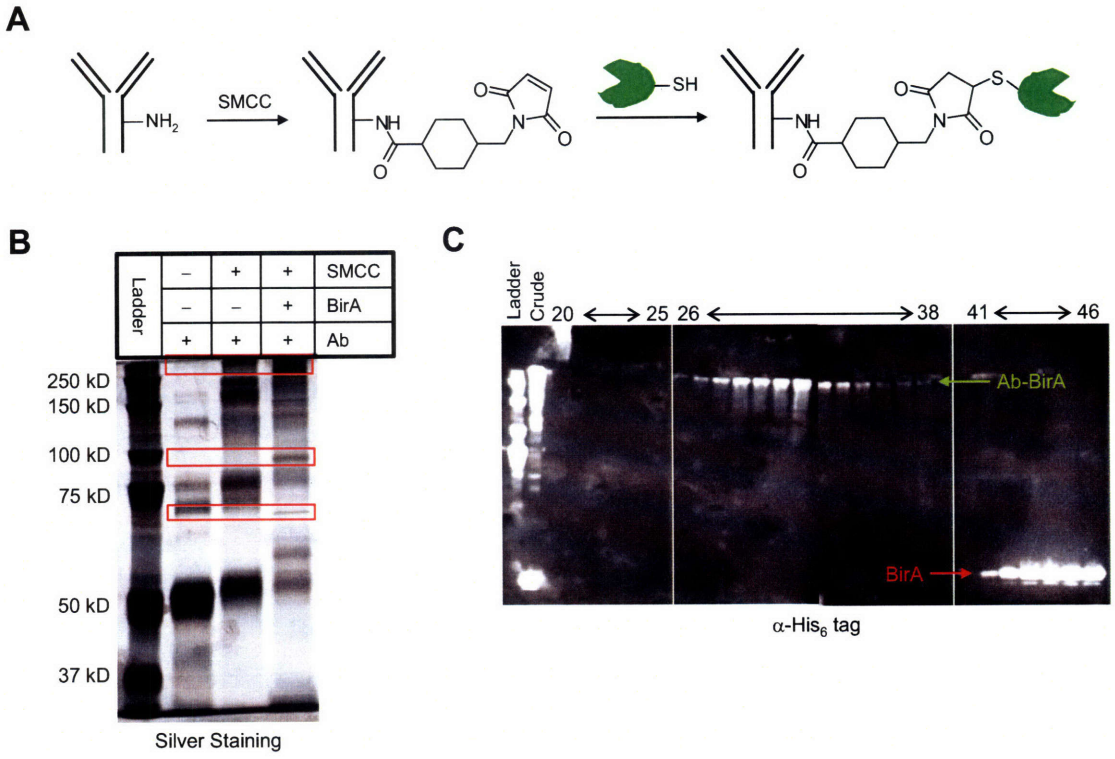


Figure 3-7: Conjugation of biotin ligase to immunoglobulins via antibody acylation by SMCC. (**A**) Conjugation scheme. Lysine residues in the antibody were acylated by the NHS-ester moiety of SMCC, and antibody-maleimide conjugates were then reacted with BirA, to form antibody-BirA conjugates. (**B**) Analysis of the conjugation products by SDS-PAGE. Anti-rabbit antibody (~ 13 μM) was reacted with 500 μM SMCC in 50 mM sodium phosphate buffer, pH 7.1, at room temperature for 1 hour. Antibody-maleimide was then combined with biotin ligase in a 1:10 ratio and incubated for 2–6 hours at room temperature. The reaction products were run on a 10% SDS-PAGE, and the extent of antibody-biotin ligase conjugation was analyzed by Silver staining. The conjugation reaction (lane 4) is compared side-by-side with a negative control with biotin ligase omitted (lanes 3). The antibody reagent before conjugation is also shown for comparison (lane 2). Red boxes highlight BirA-antibody conjugates. (**C**) Western blot analysis of elution fractions from the Superdex^{TM} 200 gel filtration column. Anti-His_6 antibody detects the C-terminal His_6 tag in biotin ligase.

First, the antibody was activated by allowing lysines to react with the NHS ester group of the SMCC heterobifunctional crosslinker and, after removing excess un-reacted SMCC, the activated antibody was conjugated to the BirA enzyme by reaction of the maleimide group in SMCC with the unique BirA surface cysteine. A schematic representation of this conjugation scheme can be seen in Figure 3-7A. This new scheme offered some interesting advantages over the previous one. There was no possibility of BirA dimerization because BirA has only one cysteine capable of reacting with the maleimide moiety. This fact eliminated the need for lengthy pre-incubations with biotin-AMP and/or NEM. Moreover, this scheme also eliminated the antibody reduction step, thus increasing the amount of antibody available for conjugation. We performed a series of iterations similar to the ones previously described to optimize the SMCC activation and final conjugation steps. Figure 3-7B confirmed that the new conjugation scheme resulted again in antibody-BirA conjugates of a defined stoichiometry. In addition to the 90–200 kD conjugates previously obtained, this scheme also resulted in BirA conjugation to the light chain of the antibody (e.g., ~ 65 kD band formed by one light chain and one BirA molecule).

Purification of antibody-biotin ligase conjugates

We tested several strategies to remove excess un-conjugated BirA and antibody from the reaction mixture. Removal of excess antibody was most important, as un-conjugated antibody would compete for antigen binding sites thus decreasing the biotinylation signal. This new conjugation scheme, however, resulted in > 90% antibody conjugation, thus eliminating the need for further purification. In contrast, the reaction mixture contained a large excess of un-conjugated BirA.

Protein G antibody pull-down followed by high-salt elution and reaction of un-conjugated BirA with biocytin-maleimide followed by pull-down with streptavidin agarose resulted in removal of up to ~ 50% of excess un-conjugated BirA. In contrast, conjugates purification by gel filtration afforded antibody-BirA conjugated with > 90% purity. We tested several gel filtration media (Sephacryl™ S200, Sephadex™ G25 and G50, and Superdex™ 200), and found that Superdex™ 200 ($1.3x10^6$ M_r exclusion limit

and 10000-600000 M_r optimal operation range) was the only one that resulted in complete separation. Figure 3-7C illustrates the separation obtained using this method. Whereas antibody-biotin ligase conjugates usually eluted in fractions 25–40, excess BirA always eluted after fraction 40–45.

Detection of antibody-biotin ligase conjugates in live cells

Before and after purification, every new batch of antibody-biotin ligase conjugates was tested for antigen binding on live cells. The protocol for live cell detection of these conjugates was similar to that previously described for the antibody-acceptor peptide conjugates. Briefly, we stained either an HA cell-surface epitope or the epidermal growth factor receptor (EGFR) with rabbit anti-HA or mouse anti-EGFR primary antibodies, followed by the anti-rabbit- or anti-mouse-biotin ligase conjugates. After washing excess antibody conjugates, bound BirA molecules (i.e., those covalently linked to the antibodies) were detected using either mouse anti-His_6 or rabbit anti-BirA, followed by the corresponding secondary antibodies conjugated to Alexa Fluor 568 or Alexa Fluor 488. Note that anti-rabbit-BirA conjugates were detected using anti-mouse-Alexa Fluor 568, whereas anti-mouse-BirA conjugates were detected using anti-rabbit-Alexa Fluor 488 in order to avoid species crossreactivity.

Figure 3-8 shows a typical result for the live cell detection of both the anti-mouse-BirA (Figure 3-8A) and anti-rabbit-BirA (Figure 3-8B) conjugates. We found that detection of the anti-mouse conjugates routinely generated a brighter signal than detection of the anti-rabbit conjugates. We believe that this is a reflection of the poor binding capacity of the anti-BirA serum compared to the mouse anti-His_6 rather than a consequence of a lower conjugation extent. This hypothesis was later confirmed on the proximity biotinylation assays (next section).

More importantly, Figure 3-8 shows that the amount of bound conjugates greatly decreases after the gel filtration purification step. We initially hypothesized that the higher signal of the un-purified conjugates could be caused by non-specific binding of un-conjugated BirA. However, a negative control where the antibody was substituted by bovine serum albumin (BSA) resulted in no detection above background, thus confirming that un-conjugated BirA was efficiently being removed from the cell-surface during the

rinses. Therefore, the decrease in signal after purification was most probably just a mere consequence of lower concentrations. Moreover, the BSA-SMCC-BirA negative control confirmed that, at least for live cell assays, there was no need for purification of the antibody-BirA conjugates.

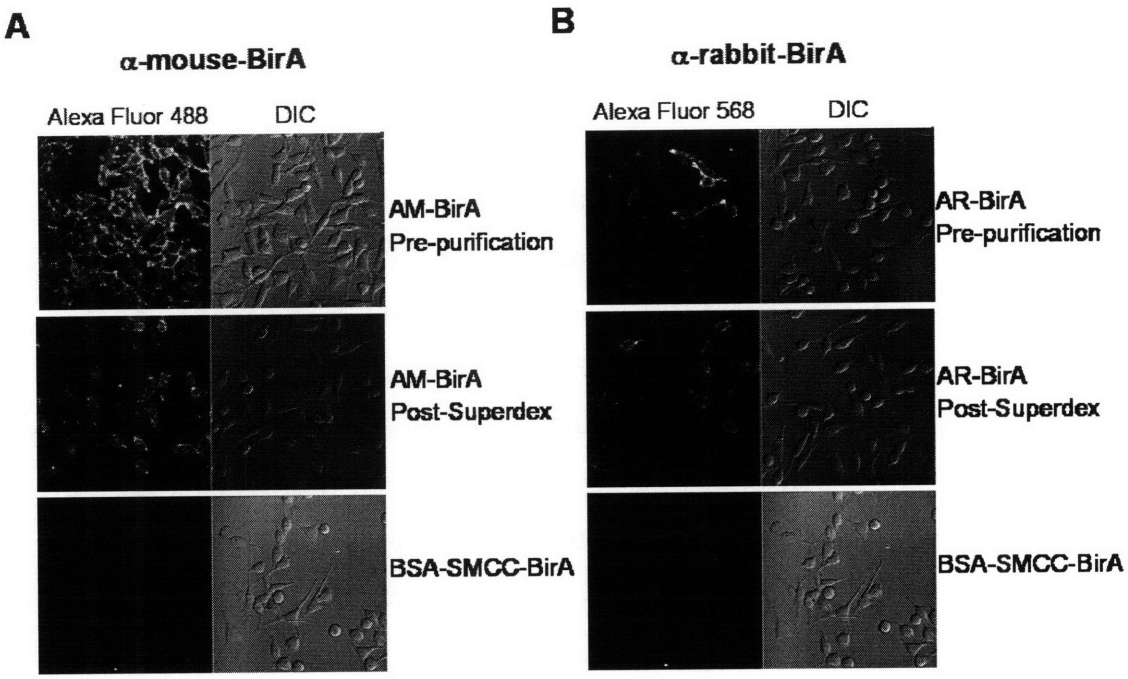


Figure 3-8: Live cell detection of antibody-biotin ligase conjugates. (**A**) Cell-surface detection of anti-mouse-BirA conjugates in live HeLa cells. Endogenous epidermal growth factor receptor (EGFR) was detected using mouse anti-EGFR, followed by the anti-mouse-BirA conjugate. After rinsing the cells to remove excess un-bound conjugates, bound BirA was detected with anti-BirA rabbit serum followed by anti-rabbit conjugated to Alexa Fluor 488. Live cell images of the introduced Alexa Fluor 568 are shown to the left of the differential interference contrast (DIC) images. (**B**) Cell-surface detection of anti-rabbit-BirA conjugates in live HeLa cells. An HA-CFP fusion was targeted to the cell surface of HeLa cells using a transmembrane (TM) domain. Cell-surface HA was detected using rabbit anti-HA, followed by the anti-rabbit-BirA conjugate. After rinsing the cells to remove excess un-bound conjugates, bound BirA was detected using mouse anti-His_6 followed by anti-mouse conjugated to Alexa Fluor 568. Live cell images of the introduced Alexa Fluor 568 are shown to the left of the DIC images. For comparison, mock conjugates where the antibody was substituted by BSA are shown as negative controls.

Antibody conjugates are capable of proximity biotinylation in live cells

We next tested the ability of the biotin ligase-antibody conjugates to biotinylate the acceptor peptide when bound to their cellular antigens. As a first approach, we

performed the proximity biotinylation on the simplest possible system: having the two antigens in the same protein, as a surrogate of an 'obligate' protein-protein interaction. For this assay, we chose HeLa cells stably transfecting a cell-surface cyan fluorescent protein (CFP), which contains an HA epitope at the N-terminus and a Myc epitope at the C-terminus. Both epitopes can be recognized by specific monoclonal antibodies, followed by the corresponding secondary antibodies conjugated to biotin ligase or the acceptor peptide, as depicted in Figure 3-9A (left). After immunodetection, excess antibody conjugates were rinsed off the cells and the biotinylation reaction was initiated by addition of 10 μM biotin and 500 μM ATP. Ligated biotin was then detected with streptavidin conjugated to Alexa Fluor 568. The live-cell images in Figure 3-9A (bottom panel) showed that transfected cells (indicated by CFP fluorescence) were labeled with Alexa Fluor 568, while neighboring untransfected cells in the same field of view were not labeled. We performed additional negative controls with omission of biotin (Figure 3-9A, top panel) or leaving out each one of the primary and secondary antibodies (data not shown), and observed no labeling in either case.

To gain confidence on the generality of the system, we performed a second assay of proximity biotinylation in live cells where the antibody conjugates were directed toward different protein molecules. We chose to detect the homodimerization of the epidermal growth factor (EGF) receptor upon stimulation with EGF (for more information on the EGF receptor, refer to next section in this chapter). The EGF receptor was recognized by a mouse monoclonal antibody against the extracellular domain, followed by a 1:1 mixture of anti-mouse-AP and anti-mouse-BirA conjugates. Biotinylation and detection was done as described above for the HA-CFP-Myc assay. The live cells images in Figure 3-9B show that cells immunostained with both conjugates were labeled with Alexa Fluor 568, whereas negative controls with biotin omitted from the labeling reaction (top left panel) or leaving out each one of the secondary antibodies (bottom panels) showed no labeling. Interestingly, a positive control where bound antibody-acceptor peptide conjugates were biotinylated by soluble exogenous BirA (middle right panel) showed a labeling signal similar to that of the proximity biotinylation, which suggests that the labeling efficiency during the proximity biotinylation assay was very high.

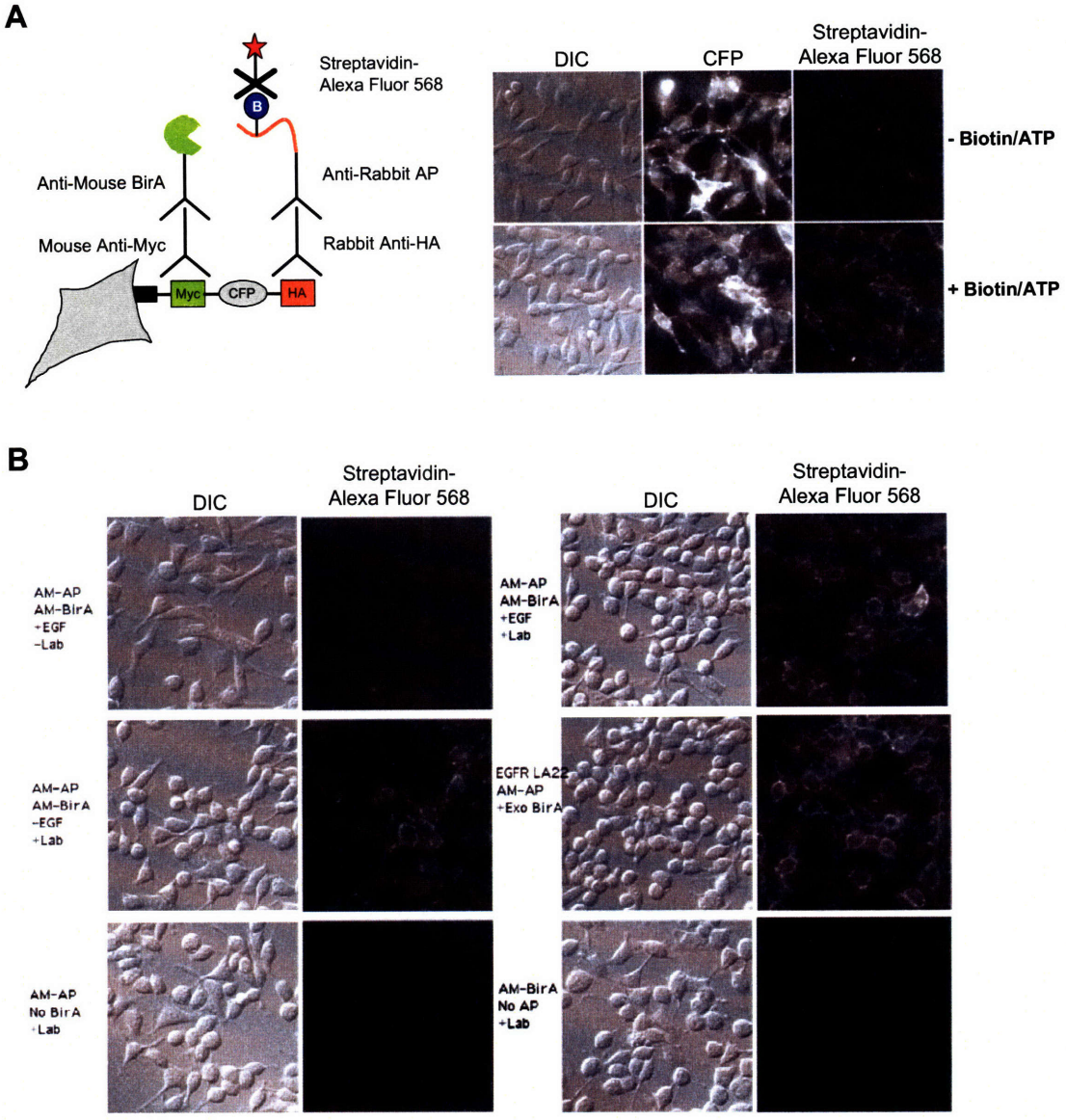


Figure 3-9: Antibody conjugates are capable of proximity biotinylation in live HeLa cells. (**A**) HeLa cells stably expressing a cell-surface HA-CFP-Myc-TM fusion protein were immunostained with mouse α-Myc and rabbit α-HA primary antibodies, followed by the corresponding α-mouse-BirA and α-rabbit-AP conjugates. Biotinylation was initiated by addition of biotin and ATP, and the introduced biotin was detected with streptavidin conjugated to Alexa Fluor 568. Live cell images of the introduced Alexa Fluor 568 are shown to the right of the CFP and DIC images, which highlight the transfected cells. A negative control with biotin omitted from the labeling reaction is shown. (**B**) HeLa cells with constitutively over-expressed endogenous epidermal growth factor (EGF) receptor were immunostained with α-EGFR mouse monoclonal antibody, followed by a 1:1 mixture of α-mouse-AP and α-mouse-BirA conjugates. Biotinylation and detection was done as in A. Labeling of cells stimulated with EGF (top right panel) is compared with un-stimulated cells (middle left panel). A positive control with BirA added in solution to fully biotinylate the anti-mouse-AP conjugates is shown in the middle right panel. Negative controls with biotin omitted from the labeling reaction (top left panel), or with either one of the antibody conjugates omitted (bottom panels) are also shown.

To our surprise, however, no difference in labeling was observed between EGF-stimulated (top right panel) and un-stimulated cells (middle left panel). We hypothesized that the presence of labeling in un-stimulated cells could have three different origins. First, HeLa cells have extremely high levels of endogenous EGF receptor expression, which could be causing a significant extent of constitutively active receptor (i.e., in dimeric form) even in the absence of EGF. Second, it is accepted that cell membranes are highly dynamic with surface receptors constantly diffusing along it. Thus, during the labeling time, the EGFR-antibody complexes could move around and occasionally 'bump' into each other, leading to a biotinylation signal. Third, because antibody binding is bivalent, immunodetection in live cells can itself lead to receptor crosslinking, and in turn to proximity biotinylation, independently of the EGF stimulus.

To test our second hypothesis, we performed the proximity biotinylation assay on two non-interacting proteins: the cell-surface HA-CFP-Myc artificial construct and the EGF receptor. We detected the HA epitope with rabbit α-HA and the EGFR with mouse α-EGFR, and we performed proximity biotinylation using our anti-rabbit-AP and anti-mouse-BirA conjugates. We observed a strong biotinylation signal that was dependent on the presence of both primary and secondary antibodies and on the addition of biotin (data not shown). These results confirmed that, indeed, the fluidic nature of the cell-membrane was allowing the antibody complexes to move around, thus generating a biotinylation signal that was not a reporter of a true protein-protein interaction. This effect was probably being enhanced in our model system because both proteins were highly over-expressed and present at extremely high concentrations at the cell-surface, thereby increasing the chances of random collision. We then decided to perform the same test on fixed cells, where we should be able to eliminate the contribution of antibody diffusion of antibody complexes as well as antibody-mediated antigen crosslinking.

Proximity bionitylation fails to detect interaction of the epidermal growth factor receptor (EGFR) with its adaptor proteins in fixed cells

Transitioning the proximity biotinylation assay into fixed samples introduced a new set of challenges; namely, the choice of an adequate fixation protocol, and the need

to remove the background signal from endogenous biotinylated proteins. We first tested several fixation protocols and chose a combination of paraformaldehyde (PFA) fixation followed by methanol permeabilization of the cell membrane. Specifically, we followed a published protocol that had been reported by Jovin and colleagues to better preserve the localization of membrane proteins such as the epidermal growth factor receptor[18]. In this protocol, the PFA fixation step ensures proper localization of membrane proteins, while the methanol permeabilization gives access to cytoplasmic interacting partners.

We next compared different fixation and blocking protocols to minimize background signal from endogenous biotinylated proteins. Mammalian cells and tissues contain biotin-dependent carboxylases, which play a key role in the synthesis of cellular metabolites during fatty acid synthesis, gluconeogenesis, lipogenesis, amino acid metabolism, and energy transduction[19]. These biotin-containing enzymes produce substantial background signals, often concentrated in the mitochondria, when biotin–avidin or biotin–streptavidin detection systems are used to identify cellular targets[20]. Figure 3-10A shows that, as expected, direct detection of endogenous biotin using sreptavidin conjugated to Alexa Fluor 568 resulted in punctuated labeling, characteristic of mitochondrial staining (left panel). We tested several blocking conditions and found that we could significantly reduce endogenous background by addition of unlabeled 100 ng/mL avidin (Figure 3-10A, right panel). We found that avidin was a better blocking reagent than streptavidin at the same concentration. This could be rationalized by the cationic nature of avidin[21], which makes it very sticky to fixed specimens. It is also worth noting that the biotin blocking protocol worked better when milk was used as the general blocking reagent, rather than the more common bovine serum albumin (BSA), and animal sera. Additionally, the surfactant 0.1% Tween 20 (polysorbate 20) was included in all steps as it was found to further reduce background staining.

An additional source of background was introduced when biotinylation was performed on fixed cells. We found that addition of biotin at concentrations higher than 10 μM resulted in increased background signal even in the absence of covalent biotin ligation by biotin ligase (Figure 3-10B). Thus, 10 μM biotin was used in all subsequent experiments of proximity biotinylation in fixed cells.

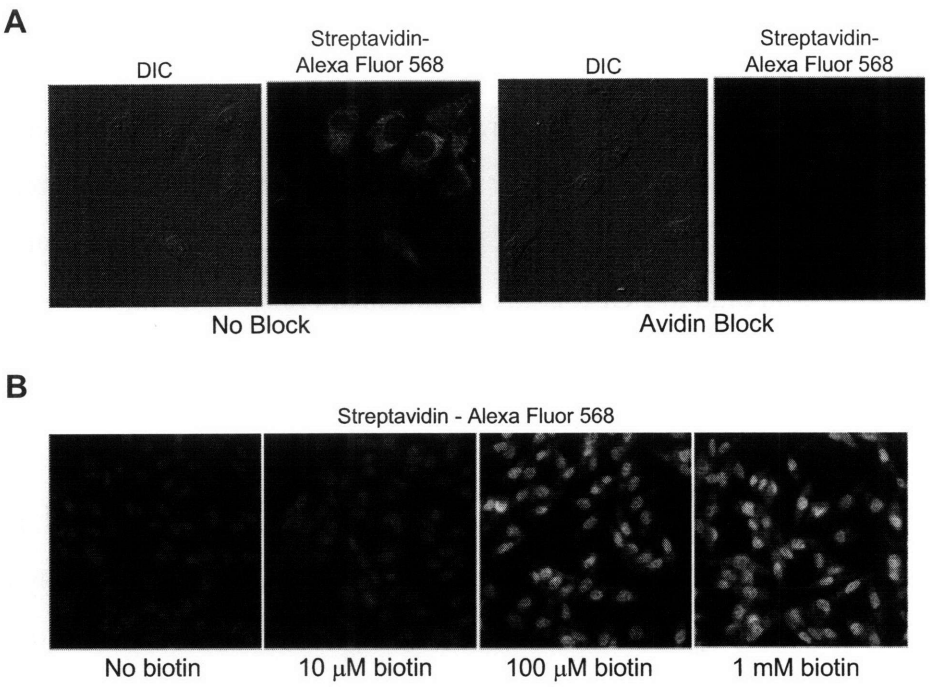


Figure 3-10: Eliminating background biotin labeling on fixed cells. (**A**) Endogenous biotin can be effectively blocked with avidin. HeLa cells were fixed with paraformaldehyde followed by methanol and blocked with 5% milk. Endogenous biotin was then blocked with 100 ng/mL avidin for 30 minutes at room temperature. After washing excess avidin, biotin was detected with streptavidin conjugated to Alexa Fluor 568. Fixed-cell images of the introduced Alexa Fluor 568 are shown to the right of the DIC image. A negative control with the avidin block omitted is also shown. (**B**) Excess biotin binds non-specifically to fixed cells. Fixed HeLa cells were first blocked with milk, followed by avidin as in A. Thereafter, biotin was added at the indicated concentrations for 1 hour at room temperature, and then washed extensively with phosphate buffered saline. Biotin was detected with streptavidin conjugated to Alexa Fluor 568.

After we optimized the fixation and blocking protocol, we first proceeded to reproduce the proximity biotinylation assay on the HA-CFP-Myc construct on fixed cells. We obtained results analogous to those of the live cell assay, albeit with a slightly lower labeling signal (data not shown). The labeling signal was probably reduced because of the need for using lower concentrations of primary and secondary antibodies to avoid non-specific antibody binding in the highly exposed fixed samples. Lower antibody concentration likely resulted in reduced antibody binding density; i.e., in live cells, the antibodies are added at high concentrations and for short periods of time, so that it is not expected that the binding reaches equilibrium, as it is for fixed-cell immunodetection.

We were then ready to test our protein-protein interaction scheme on a real pair of interacting proteins. For our initial proof-of-principle, we chose the EGF receptor interaction with some of its adaptor proteins. The epidermal growth factor receptor

(EGFR, also known as ErbB1) is the archetypal member of a family of four receptor tyrosine kinases, which also include ErbB2, ErbB3 and ErbB4. These receptors are responsible for mediating several cellular responses such as cell proliferation and differentiation, and an imbalance in one or more of these receptors has been implicated in several epithelial carcinomas[22]. Epidermal growth factor (EGF) and transforming growth factor α (TGFα) both bind to EGFR at the plasma membrane. After ligand binding, EGFR dimerizes, its intrinsic tyrosine kinase is activated, and specific tyrosine residues in its cytoplasmic tail become auto-phosphorylated[23]. These phosphotyrosine residues then serve as docking sites for the so-called adaptor proteins, which are signal transducers containing Src homology 2 (SH2) or phosphotyrosine binding (PTB) motifs[24]. Binding of the adaptor proteins to EGFR initiates many signalling cascades that ultimately result in a physiological outcome. Figure 3-11 schematically represents the EGFR with its tyrosine sites of auto-phosphorylation as well as the tyrosines phosphorylated by the Src kinase. Among the several signal transducing pathways activated by EGF binding, the best characterized is the Ras-Raf-MEK-MAPK pathway, which is initiated by Grb2-Sos or Shc-Grb2-Sos association. In addition, Ras-specific GTPase-activating protein (Gab), ubiquitin ligase (Cbl), phosphatidyl inositol 3-kinase (PI3K), and phospholipase C γ (PLCγ) signaling pathways are initiated from activated EGFR[24]. Finally, ligand-bound EGFR is rapidly internalized through receptor-mediated endocytosis and sorted to multivesicular bodies. The receptor either recirculates to the plasma membrane after ligand dissociation or is degraded through the lysosomal pathway.

We chose the EGFR interaction with its adaptor proteins because it is a well-characterized system and because some of these interactions have been previously studied by co-localization analysis of classic immunofluorescence images. However, many of these studies reported contradictory results. For example, the localization of the adaptor protein Shc before EGF stimulation was reported predominantly cytoplasmic by Oksvold and colleagues[25], whereas other groups claimed that un-phosphorylated Shc was mainly located in the endoplasmic reticulum[26,27] of the same A431 cells. Consistently, the reported extent of co-localization between EGFR and Shc at the plasma membrane upon addition of EGF is also highly variable, ranging from 5–15%[28]. In most cases, the differences in the localization pattern of the whole population of an adaptor protein

before and after stimulation were so small that no quantification was reported[25,29,30]. We hypothesized that these inconsistencies could in part be caused by the technical difficulties of the co-localization analyses and the intrinsically limited spatial resolution of the immunofluorescence technique.

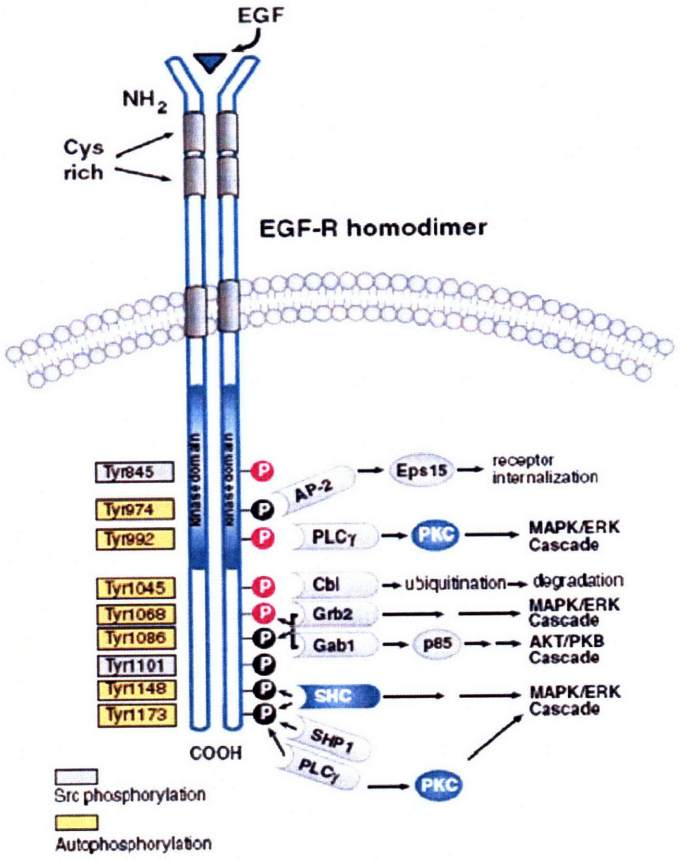


Figure 3-11: Schematic representation of the epidermal growth factor receptor. The epidermal growth factor (EGF) receptor is a ~ 180 kD cell-surface protein with an extracellular N-terminal portion containing the ligand binding and the dimerization domains, and a C-terminal portion, which includes the kinase domain and the cytoplasmic tail, where the adaptor proteins bind. Tyrosine residues auto-phosphorylated by the EGF receptor kinase domain are marked in yellow, whereas Tyrosine residues phosphorylated by the Src kinase are marked in grey. Some of the adaptor proteins binding to the different phosphorylated tyrosines are indicated, together with the signaling pathways that they initiate. AP-2: adaptor protein 2. Eps15: EGFR pathway substrate 15. PLCγ: phospholipase Cγ. PKC: protein kinase C. MAPK: mitogen-activated protein kinase. ERK: extracellular signal-regulated kinase. Cbl: ubiquitin ligase. Grb2: growth factor receptor-bound protein 2. Gab1: Ras-specific GTPase-activating protein 1. Shc: Src homologus and collagen-like protein. Shp1: Src homology phosphatase 1. Figure taken from the Santa Cruz Biotechnology website.

We first performed classical immunofluorescence detection of the EGF receptor and the adaptor proteins. We routinely worked with three different cell lines – HeLa, A431, and NR6 – because of their high EGFR expression levels. HeLa and A431 cells are human epithelial carcinoma cell lines constitutively expressing 10^4 EGF receptors/cell and 10^6 EGF receptors/cell, respectively[31,32]. NR6-WT cells were derived from the EGFR-negative variant of the 3T3 mouse fibroblasts cell-line (NR6 parental)[33], which were later transfected with over-expressed human receptor (~ 10^4 receptors/cell)[34]. In our hands, the NR6 cell-line was the most responsive to EGF stimulation, however, because most of our antibodies against the adaptor proteins were specific for human, this cell-line was only useful for studies of EGFR homodimerization (see below). In contrast, the A431 cells worked best in terms of adaptor protein recognition, but the effect of EGF stimulation on EGFR activation was highly variable. As it can be seen in Figure 3-12, the basal levels of phosphorylated EGFR (pEGFR) were relatively high even in un-stimulated cells, probably due to constitutive dimerization caused by receptor over-expression. Figure 3-12 also depicts the different localization patterns of the adaptor proteins before and after stimulation with 100 ng/mL EGF for 10 minutes at room temperature. As expected, we observed that all the adaptor proteins except CrkII changed their localization upon EGF stimulation, moving towards the plasma membrane. However, the extent of the change varied for the different proteins. The most dramatic differences were observed for Src kinase and the Grb2 adaptor protein, which not only changed localization, but also showed a marked increased in the overall detected signal. On the other hand, whereas a certain membrane population of Shc and Nck was observed upon EGF stimulation, the majority of the signal still remained reticular and cytosolic, respectively.

Co-localization analysis of several cells and samples from different days revealed that Grb2 exhibited the largest increase in the co-localization extent with EGFR, changing from 47% ± 3% to 86% ± 5% upon EGF stimulation. Co-localization of EGFR and Shc also changed significantly, varying from 41% ± 3% before EGF stimulation to 68% ± 3% after 10 minutes incubation with 100 ng/mL EGF. Co-localization percentages for Src and Nck were too variable among experiments to extract meaningful conclusions.

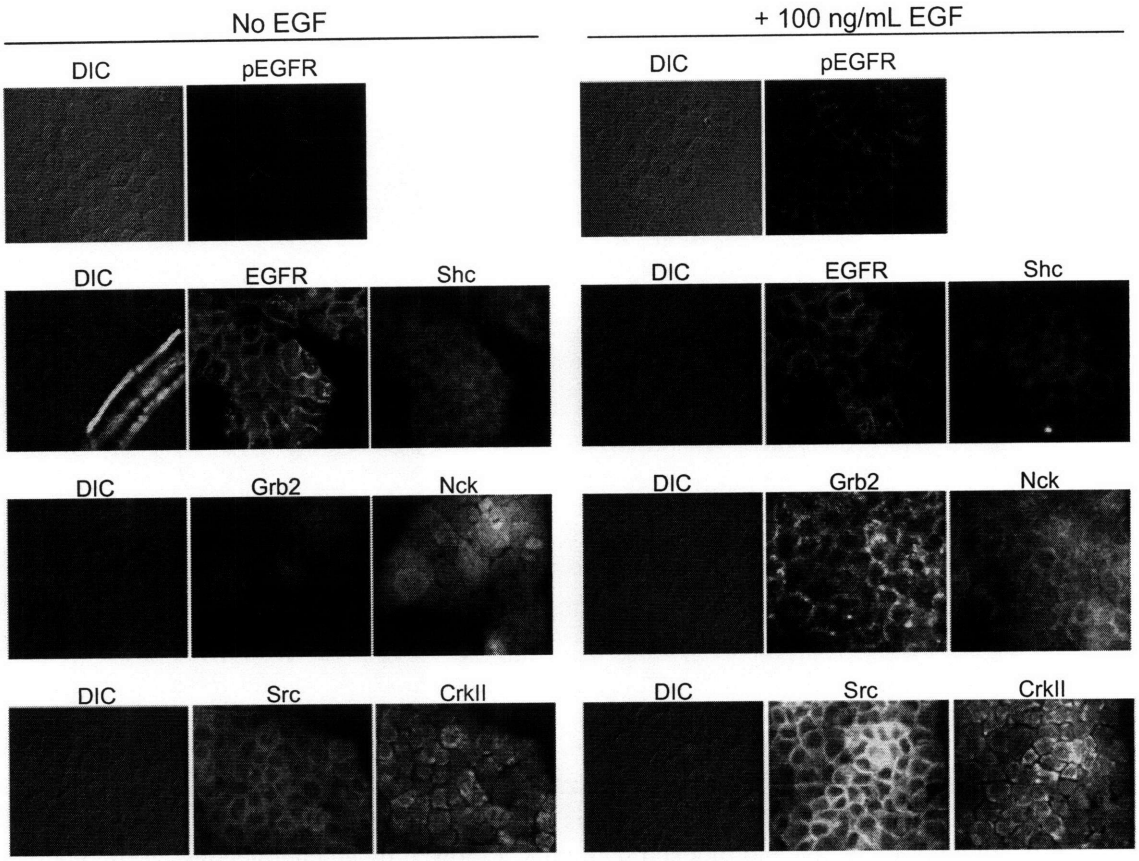


Figure 3-12: Immunofluorescence detection of the epidermal growth factor (EGF) receptor and its adaptor proteins in A431 cells. The localization of the EGF receptor, the Src kinase, and the adaptor proteins Shc, Grb2, Nck, and CrkII is shown before and after stimulation with 100 ng/mL EGF for 10 minutes at room temperature. Additionally, EGF receptor activation was confirmed with an anti-phosphotyrosine antibody (pEGFR). Each antigen was detected with a specific primary antibody (see Experimental section for details) followed by the corresponding anti-mouse or anti-rabbit conjugated to Alexa Fluor 568 or Alexa Fluor 488, respectively. The Alexa Fluor images (AF-568 middle panels, AF-488 right panels) are shown to the right of the DIC images.

Next, we used our biotin ligase and acceptor peptide antibody conjugates to study the EGF-induced homodimerization of EGFR. We performed the assay as described above for the live-cell tests (see Figure 3-9B), except that we fixed and permeabilized the cells prior to immunolabeling. A431 cells were fixed, blocked with milk, and endogenous biotin was blocked with avidin. The EGF receptor was then recognized by a mouse monoclonal antibody against the extracellular domain, followed by a 1:1 mixture of anti-mouse-AP and anti-mouse-BirA conjugates. Biotinylation was initiated by addition of 10 μM biotin and 500 μM ATP to the fixed cells, and ligated biotin was detected with streptavidin conjugated to Alexa Fluor 568. Figure 3-13 shows the results.

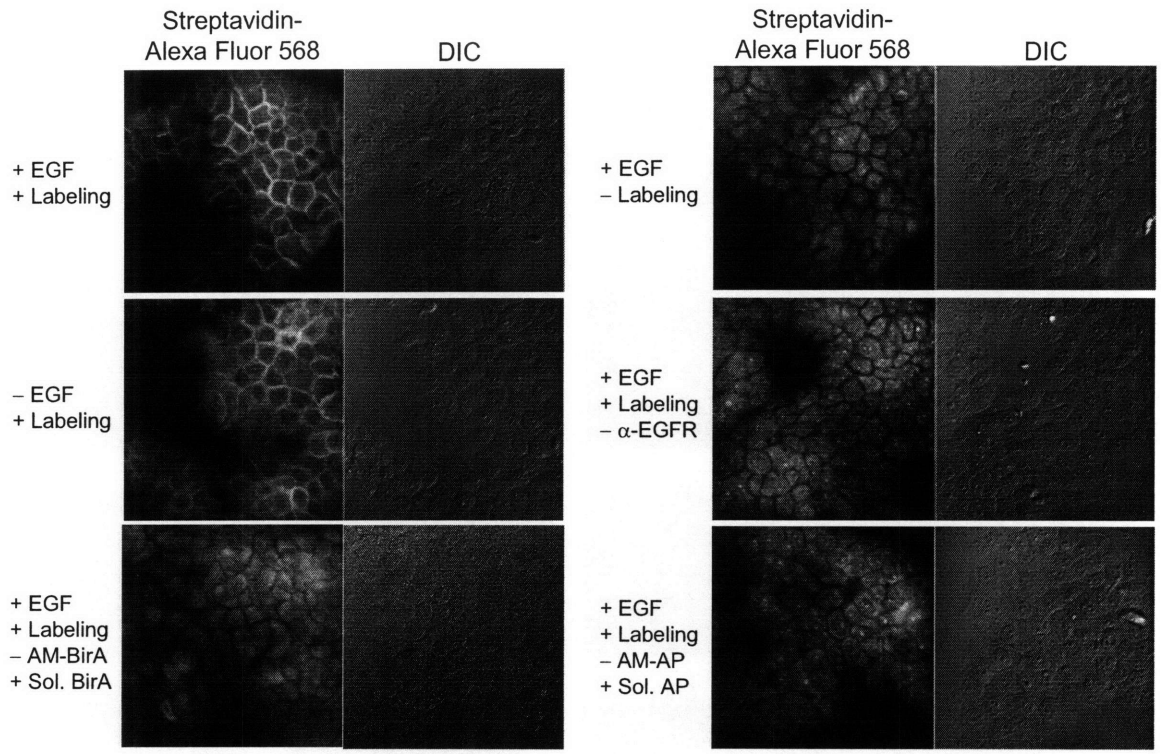


Figure 3-13: Detection of EGFR homodimerization in fixed A431 cells by proximity biotinylation. Epidermoid carcinoma A431 cells were fixed with paraformaldehyde (PFA), permeabilized with methanol, and blocked with milk, followed by an endogenous-biotin blocking step with avidin. Thereafter, cells were immunostained with α-EGFR mouse monoclonal antibody, followed by a 1:1 mixture of α-mouse-AP and α-mouse-BirA conjugates. Biotinylation was initiated by addition of biotin and ATP, and the introduced biotin was detected with streptavidin conjugated to Alexa Fluor 568. Fixed-cell images of the introduced Alexa Fluor 568 are shown to the left of the DIC images. Labeling of cells stimulated with EGF (top left panel) is compared with un-stimulated cells (middle left). Negative controls with biotin omitted from the labeling reaction (top right), or without each one of the primary or secondary antibodies are also shown. In the case where the secondary antibodies were omitted, they were substituted by soluble un-conjugated biotin ligase (bottom left) or acceptor peptide (bottom right), respectively.

We observed that A431 cells immunostained with both antibody conjugates exhibited a clear membrane labeling pattern in the Alexa Fluor 568 channel, whereas negative controls with biotin omitted from the labeling reaction (top right images) or leaving out each of the primary or secondary antibodies (middle right, and bottom left and right images) showed no significant labeling signal above background. It must be noted that the signal/background ratio observed was overall very low, thus enhancing the background labeling caused by streptavidin non-specific or specific-binding to endogenous biotinylated proteins. Unfortunately, as in the live-cell assay, we observed no difference in labeling between EGF-stimulated (top left) and un-stimulated cells (middle

left). These results contradicted our previous hypothesis, which stated that the lack of EGF-induced increase in biotinylation signal was caused by random-biotinylation during dynamic receptor movement within the membrane of living cells. However, we had previously found that the response to EGF in A431 cells was highly variable, as detected by the anti-phosphotyrosine antibody, owing to the high-levels of EGFR over-expression (10^6 receptors/cell). Therefore, in order to rule out the possibility that the lack of difference was caused by absence of EGF-induced EGFR homodimerization, we decided to repeat the proximity biotinylation test on NR6, a cell-line that had routinely shown a far more pronounced EGF-induced increase in EGFR phosphorylation than that observed in A431 cells.

To our discontent, we found that the results were the same as we had observed for A431 cells (Figure 3-14). Although the signal was a lot higher for NR6 than for A431 cells, we still observed no difference in the labeling extent of EGF-treated and un-stimulated cells. These results suggested that the biotinylation signal we were observing was not reporting on a true protein-protein interaction – the homodimerization of EGFR – but it was rather just an artifact of antibody-conjugates, which, being polyclonal, were likely binding to the same primary α-EGFR antibody. We tested different EGFR antibodies in an attempt to eliminate this multivalent binding, but we obtained similar results with all of them (data not shown). We also observed no difference when we attempted to enhance the effect of EGF stimulation by co-addition of the phosphatase inhibitor vanadate, which should result in an overall increase in the number of phophorylated receptors (data not shown).

Finally, because the artifact was restricted to homodimerization studies, we sought to eliminate it by applying the proximity biotinylation scheme to the interaction of EGFR with one of its adaptor proteins. We focused mainly on the Grb2 and Shc adaptor proteins because they have shown biggest changes in their co-localization with EGFR upon EGF stimulation (see above). We tested several combinations of α-EGFR and α-Shc or α-Grb2 antibodies, as well as our three main cell lines (HeLa, NR6-WT, and A431) and various conditions for EGF stimulation, and had no success. We never observed biotinylation signal when the antibody-acceptor peptide and antibody-biotin ligase conjugates were targeted to different proteins.

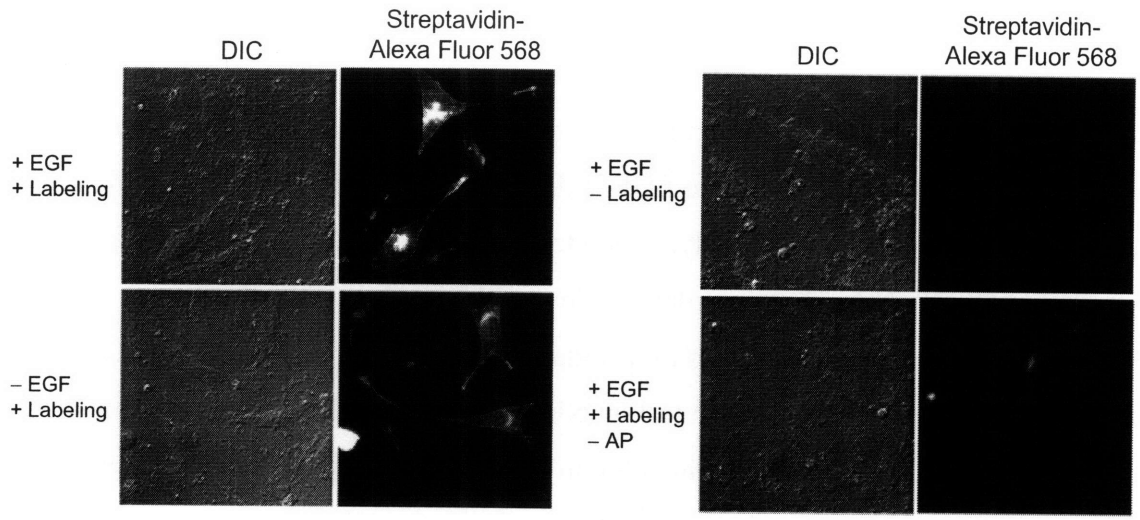


Figure 3-14: Detection of EGFR homodimerization in fixed NR6 cells by proximity biotinylation. Mouse fibroblasts over-expressing human EGF receptor (NR6-WT) were fixed with PFA, permeabilized with methanol, and blocked with milk, followed by an endogenous-biotin blocking step with avidin. Thereafter, cells were immunostained with α-EGFR mouse monoclonal antibody, followed by a 1:1 mixture of α-mouse-AP and α-mouse-BirA conjugates. Biotinylation was initiated by addition of biotin and ATP, and the introduced biotin was detected with streptavidin conjugated to Alexa Fluor 568. Fixed-cell images of the introduced Alexa Fluor 568 are shown to the right of the DIC images. Labeling of cells stimulated with EGF (top left images) is compared with un-stimulated cells (bottom left). Negative controls with biotin omitted from the labeling reaction (top right), or without anti-mouse-acceptor peptide conjugate (bottom right) are also shown.

Discussion

While we still think that the concept of proximity biotinylation is powerful, our experience illustrates that the system requires positioning of biotin ligase and the acceptor peptide in close enough proximity and in a productive conformation, which is hard to achieve when the enzyme and peptide are both linked by chemical conjugation to large antibody molecules. As we describe in Chapter 4 however, we later demonstrated that the concept of proximity biotinylation is indeed valid and can be successfully applied to the detection of protein-protein interactions if the appropriate flexibility is provided to BirA and AP.

We believe that three main factors prevented proximity biotinylation of AP by BirA in the context of IF. First, the large size of an antibody (~ 150 kD) was not only likely to hinder the acceptor peptide (~ 2 kD), but could also potentially separate BirA

and AP further than what they could reach. That is, the separation between biotin ligase and its acceptor peptide within the context of the antibody conjugates was far greater than the distance intrinsically defined by the protein-protein interaction. Although this problem is fundamental to the methodology, we feel that it could have been circumvented, had we implemented better conjugation strategies. Second, the overall yield of the conjugation reaction between biotin ligase the antibody was very low, which resulted in inefficient separation from excess un-reacted antibody and biotin ligase molecules. Third, and perhaps most important, factor was thus the choice of chemical conjugation strategies. On the one hand, we conjugated the AP to the antibody directly at the peptide C-terminus. Although BirA has been shown to recognize AP when fused to the N-terminus, C-terminus, or an internal loop[35], AP is, in these contexts, within a flexible protein structure and protruding towards the exterior of the protein. This is probably not the case in our antibody-acceptor peptide conjugates, where AP was conjugated via the antibody sugars. Examination of the crystal structure of the crystallizable portion (Fc) of an immunoglobulin G (Figure 3-15A) reveals that the carbohydrates attached to the heavy chains lie at the interface between the two heavy chains, and are thus not highly exposed to the protein exterior. We reason that conjugation of the acceptor peptide via free hinge-region thiols, or with addition of a ~ 10-amino acid linker between the hydrazide functional group and the 15-amino acid recognition sequence, would have probably improved AP accessibility. On the other hand, biotin ligase was chemically conjugated to the antibody via its unique Cysteine 107 and using SMCC as the only spacer (~ 8 Å). Figure 3-15B illustrates the position of Cys107 relative to the active site. It is clear that the side chain of Cys107 is extremely close to the active site (~ 8 Å calculated distance from the sulfur atom in Cys107 to one of the ureido nitrogens in biotin). Although Cys107 is not essential for catalysis, and we showed that SMCC conjugation did not affect enzymatic activity, it is likely that attachment to the antibody via Cys107 holds the enzyme in a rather rigid position, with its active site very close to the antibody. We propose that a better conjugation strategy could involve mutating Cys107 to a serine, and introducing an extra cysteine at the Gly55 position, which is positioned at a β-turn within the enzyme N-terminal binding domain (highlighted green in Figure 3-15B). The N-terminal domain is only used for DNA

binding and it is joined to the C-terminal catalytic domain by a flexible linker. Therefore, linking BirA to the antibody via its N-terminal domain could perhaps provide the enzyme with enough freedom to productively recognize the acceptor peptide.

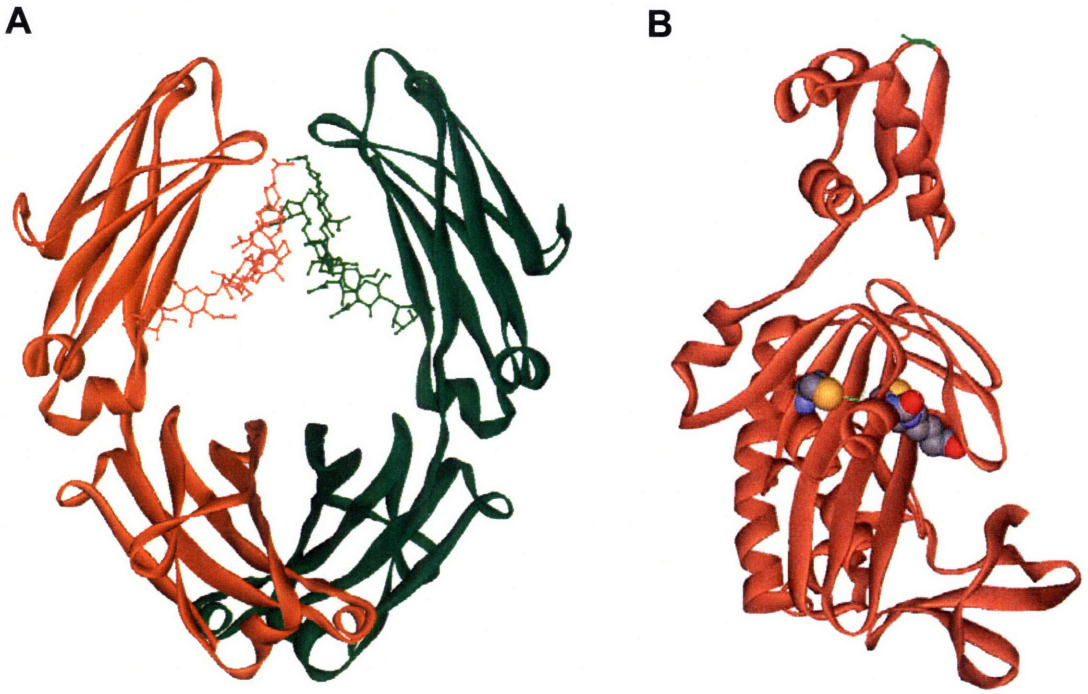


Figure 3-15: Ribbon representations of the Fc of human immunoglobulin G and biotin ligase. (**A**) Crystal structure of the crystallizable portion (Fc) of human immunoglobulin G (PDB: 1H3Y). The two heavy chains are colored green and orange. Amino acid residues are represented in ribbon format, whereas carbohydrate groups are represented as ball and stick. (**B**) Position of biotin ligase Cysteine 107 with respect to the active site. The entire amino acid sequence of *E. coli* biotin ligase in the crystal structure (PDB: 1HXD) is shown as a solid ribbon. Cysteine 107 (left) and biotin (right) are shown in CPK format. Carbon atoms are colored grey, oxygen atoms are colored red, nitrogen atoms are colored blue, and sulfur atoms are colored yellow. These images were generated from the original PDB files using DS Viewer Pro software.

In the last months of development of this project, a new method to detect endogenous protein-protein interactions was reported that employed an scheme analogous to ours[36]. In the proximity ligation technique, each antibody is conjugated to a single-stranded DNA molecule. When two or more antibodies bind to either the same molecule (e.g., to detect the presence of a protein) or to two different, interacting molecules, the DNA strands are brought into proximity and, after addition of a hybridizing connector oligonucleotide, they can be joined by ligation. The newly formed chimeric DNA strand can then be detected and quantified by real-time PCR with

extremely high sensitivity[37]. For the detection of PPIs in fixed cells, a combination of the proximity ligation technique with rolling-circle amplification resulted in detection of endogenous Myc-Max complexes with single-molecule spatial resolution[36]. Rolling-circle amplification was used in place of PCR to avoid thermal cycling and to ensure that the amplification product remains attached to the antigen. Proximity ligation has most recently also been applied to the detection of post-translational modifications, such as phosphorylation of the platelet-derived growth factor receptor[38].

The proximity ligation technique is in essence the same as our proposed proximity biotinylation except that the enzyme-substrate pair is substituted by DNA strands. These DNA strands are typically 30–35 nucleotides long, and can thus expand much longer distances that our enzyme-substrate pair. Additionally, the sensitivity of the proximity ligation method is expected to be higher than what could be obtained using proximity biotinylation owing to the exceptional signal amplification of the rolling circle amplification.

Conclusions

In this chapter, we have described our efforts to develop a new methodology for the detection of endogenous protein-protein interactions. With this method, we sought to bridge the gap between the immunofluorescence (IF) and FRET studies to be able to report, for the first time, on true protein-protein interactions between endogenous proteins within the context of fixed cells. The proposed methodology was an adaptation of the IF technique, where two antibodies were conjugated to biotin ligase and its acceptor peptide instead of fluorophores. Upon protein-protein interaction, the enzyme and its substrate peptide should be brought into close proximity, where biotinylation could occur. We first developed robust protocols for chemical conjugation of immunoglobulins to the acceptor peptide and biotin ligase. We then validated the binding capacity and the catalytic activity of the conjugates both *in vitro* and in live cells. The antibody conjugates failed however to detect the interaction between the epidermal growth factor receptor and some of its adaptor proteins in fixed cells. We hypothesized that the inability of our system to detect

PPIs in fixed cells was caused by the intrinsically unfavorable geometry imposed by the antibodies on BirA and AP, by our choice of strategies for chemical conjugation, and by the overall low yield of the bioconjugation reactions. Finally, a similar method called proximity ligation was later reported by the laboratory of Ulf Landegren (U. of Upsala, Sweden), which overcame many of the technical challenges encountered in the proximity biotinylation scheme. Proximity ligation has successfully been applied to the detection of interactions between the Myc and Max transcription factors in fixed mammalian cells[36]. It remains to be seen if the method finds widespread application to other protein pairs and by other research groups.

Experimental

Bacterial expression and purification of *E. coli* biotin ligase

The pBTac2-BirA plasmid for expression of *E. coli* biotin ligase was a gift of Prof. Beckett (University of Maryland). To enhance expression in *E. coli* BL21(DE3), the b*irA* gene was subcloned into the *EcoRI* and *HindIII* restriction sites of pET21a for expression of a C-terminally His_6-tagged enzyme. The pET21a-BirA plasmid was transformed into *E. coli* BL21(DE3) cells, which were amplified in LB media supplemented with 100 μg/mL ampicillin at 37 °C until OD_{600} 0.9. Protein expression was induced with 100 μg/mL IPTG for 3 hours at 30 °C. Thereafter, cells were harvested by centrifugation (6,000 rpm, 10 minutes, 4 °C) and the pellet was resuspended in lysis buffer (50 mM Tris base, 300 mM NaCl, pH 7.8) containing 2.5 mM phenylmethylsulfonyl fluoride (PMSF) and protease inhibitor cocktail (Calbiochem). Cells were lysed by ultrasonic treatment (six 15-second bursts, with 1 minute of cooling to 4 °C between bursts). The extract was cleared by centrifugation (13,200 rpm, 10 minutes, 4 °C) and the His_6-tagged protein was purified using Ni-NTA agarose (Qiagen). Fractions were analyzed by 10% SDS-PAGE followed by Coomassie staining. Fractions containing the enzyme were pooled and dialyzed against PBS. Protein concentrations were measured using the BCA assay (Pierce) with BSA as the reference

standard. For each new enzyme preparation, catalytic activity was confirmed by running a standard HPLC activity assay (see below).

HPLC assay of biotin ligase activity

The synthetic acceptor peptide (AP) with sequence KKKGPGGLNDIFEAQKIE WH was synthesized by the Tufts University Core Facility. The reaction conditions to test the specific activity of biotin ligase were as follows: 50 mM bicine pH 8.3, 5 mM magnesium acetate, 4 mM ATP, 100 μM AP, 200 nM BirA, and 1 mM biotin. The reaction was incubated at 30 °C for 20 minutes before quenching with the addition of 45 mM ethylenediamine tetraacetic acid (EDTA). Reactions were analyzed on a reverse-phase HPLC column (Microsorb-MV 300 C18) using a gradient of 10-43% acetonitrile/water with 0.1% TFA over 20 minutes (flow rate 1.0 mL/min). Retention times were 8.2 minutes for biotin, 16.3 minutes for the AP, and 17.7 minutes for the AP-biotin conjugate. Under the above reactions conditions, ~ 50% of the AP should be converted to AP-biotin (k_{cat} = 12 min^{-1})[4].

Sodium periodate oxidation of IgG molecules

Antibody solutions (typically 100 μL at 1–2 mg/mL) were reacted with 1 mM sodium periodate for 30 minutes at room temperature in PBS, pH 7.2. The reactions were kept in the dark, and the excess sodium periodate was immediately removed by gel filtration using a Sephadex™ G25 resin (Quick Spin or Nap5 columns, GE Healthcare). Oxidized-antibody concentrations were calculated from the absorbance at 280 nm using the extinction coefficient of 1.4 L g^{-1} cm^{-1}. Typically, affinity purified goat anti-rabbit or anti-mouse IgGs (Rockland Immunochemicals) were used in the conjugations.

Conjugation of the acceptor peptide to oxidized antibodies

A synthetic N-terminally hydrazide-functionalized AP was synthesized by the Tufts University Core Facility. The crude peptide was purified by reverse-phase HPLC (Microsorb-MV 300 C18, 10-39% acetonitrile/water with 0.1% trifluoroacetic acid

(TFA) over 35 minutes, flow rate 4.7 mL/min); the desired peak had a retention time of 28 minutes. Following lyophilization, the peptide was redissolved in water, and the concentration was determined from the absorbance at 280 nm using the calculated extinction coefficient of 5690 $M^{-1}cm^{-1}$.

Purified AP-hydroxylamine was added to the oxidized antibody solution (~ 100 µL, see above) to a final concentration of 1 mM, and the reaction was allowed to proceed overnight at room temperature in 10 mM sodium acetate, pH 5. Thereafter, the formed hydrazone was reduced by reaction with 100 mM sodium cyanoborohydride ($NaBH_3CN$) for 6 hours at room temperature, with an additional supplement of $NaBH_3CN$ after the first 3 hours. Finally, un-reacted aldehydes in the oxidized antibody were blocked by reaction with 50 mM ethanolamine, pH 9.6 for 30 minutes at room temperature. The reaction mixture was then dialyzed against PBS pH 7.4 and purified as described below.

Affinity purification of antibody-acceptor peptide conjugates

Dialyzed antibody-acceptor peptide mixtures were incubated with Protein G (Calbiochem) resin for 2 hours at room temperature in the presence of 3% BSA. Prior to sample loading, the resin was pre-blocked by incubation in 3% BSA in PBS-T for 15–30 minutes at 37 °C. The amount of resin required for the purification was calculated based on a binding capacity of 1–5 mg of goat IgGs per 1 mL of resin (info provided by the manufacturer). The antibody was eluted from the resin with 100 mM glycine pH 3.0, and the elution fractions were immediately neutralized to pH 7.0 with 500 mM Tris base, and dialyzed against PBS. To determine the yield of the conjugation procedure, the purified mixtures were incubated with 5 µM biotin ligase, 1 mM biotin, 4 mM ATP, and 5 mM magnesium chloride in 50 mM bicine buffer, pH 8.3, and the extent of AP biotinylation was analyzed by Western blotting as described below.

Conjugation of biotin ligase to oxidized antibodies by reductive amination

Oxidized antibody and purified biotin ligase were first separately dialyzed against 200 mM sodium carbonate/bicarbonate pH 9.6. Thereafter, the solutions were mixed to

obtain a 4:1 to 15:1 ratio of BirA to antibody and the reaction was allowed to proceed for 2 hours at room temperature. The formed imines were then reduced with 50 mM $NaBH_3CN$ for 30 minutes at room temperature and the unreacted aldehydes were blocked by incubation with 50 mM ethanolamine for an additional 30 minutes. Excess reagents were removed by gel filtration (G25 Quick Spin columns, GE Healthcare) or by dialysis against PBS, pH 7.4, and the reaction products were analyzed by Western blotting as described below.

Synthesis of adipic dihydrazide-biotin ligase conjugates

In a 27 µL reaction, 40 µM biotin ligase was combined with 5 mM 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride (EDC) in 100 mM sodium carbonate pH 8.1. After 30 minutes at room temperature, adipic dihydrazide (3 µL, 500mM) was added and the reaction was allowed to proceed for 4 hours at room temperature. The protein was purified from excess EDC, sulfo-NHS, and adipic dihydrazide using size-exclusion chromatography on Sephadex™ G-25 (Nap5 column, GE Healthcare).

Conjugation of biotin ligase-hydrazide to oxidized antibodies

In a 20 µL reaction, 12.5 µL of biotin ligase-hydrazide (C_f = 10 µM) were combined with 2 µL of oxidized antibody (C_f ~ 1 µM) in 50 mM 2-(N-morpholino) ethanesulfonic acid (MES) pH 6.2 and incubated at room temperature for 2 hours. Thereafter, the mixture was cooled down to 4 °C and the hydrazone was reduced by addition of 50 mM sodium cyanoborohydride ($NaBH_3CN$) for 30 minutes at room temperature or 4 °C. Unreacted aldehyde groups were then blocked with 50 mM ethanolamine pH 9.6 for 30 minutes at room temperature. The final reaction mixture was dialyzed against PBS, pH 7.4, and analyzed by Western blotting as described below.

Synthesis of SMCC-biotin ligase conjugates

Biotin ligase was reacted with succinimidyl 4-[N-maleimidomethyl] cyclohexane-1-carboxylate (SMCC). Biotin ligase was first pre-incubated for 5 minutes at room

temperature with 100 μM biotin and 400 μM ATP, followed by 30 minutes incubation with 1 mM N-ethylmaleimide (NEM), before addition of SMCC. Thereafter, SMCC was added to 50 μL of BirA solution (~ 20–30 μM) to a final concentration of 200 μM in 50 mM sodium phosphate buffer, pH 7.1 and the reaction was incubated at room temperature for 1 hour. The biotin ligase-SMCC conjugates were purified from excess SMCC using size-exclusion chromatography on a NAP-5 (Sephadex™ G-25 resin, GE Healthcare) column. The proteins were eluted from the column in 7×100 μL fractions in PBS, pH 7.2. The most concentrated fractions (usually fractions 3–5) were combined and immediately reacted with the reduced antibody (see below).

Antibody reduction to expose free thiols

In a 100 μL reaction, 98 μL of antibody solution (2 mg/mL) were combined with 2 μL of of 1M dithiothreitol (DTT, C_f = 20 mM) in PBS pH 7.2 and incubated at room temperature for 30 minutes. Thereafter, excess DTT was quickly removed using size-exclusion chromatography on a NAP-5 (Sephadex™ G-25 resin, GE Healthcare) column. The antibody was eluted from the column in 6×100 μL fractions in PBS, pH 7.2 (note that using larger elution volumes is not recommended as they may elute DTT). The most concentrated fractions of reduced antibody (usually fractions 3–5) were combined and immediately combined with the BirA-SMCC solution. Typically, affinity purified goat anti-rabbit or anti-mouse IgGs (Rockland Immunochemicals) were used for these conjugations.

Synthesis of SMCC-antibody conjugates

Goat affinity purified anti-mouse or anti-rabbit IgGs (Rockland Immunochemicals) were reacted with succinimidyl 4-[N-maleimidomethyl] cyclohexane-1-carboxylate (SMCC). SMCC was added to 50 μL of antibody solution (2 mg/mL) to a final concentration of 500 μM in 50 mM sodium phosphate buffer, pH 7.1, and the reaction was allowed to proceed at room temperature for 1 hour. The antibody-SMCC conjugates were purified from excess SMCC using size-exclusion chromatography on a NAP-5 (Sephadex™ G-25 resin, GE Healthcare) column. The proteins were eluted from

the column in 7×100 μL fractions in PBS, pH 7.2. The most concentrated fractions of antibody-SMCC (usually fractions 3–5) were combined and immediately reacted with BirA.

Conjugation of biotin ligase-SMCC-antibodies

For the first method of conjugation, BirA-SMCC (~ 300 μL) was combined with reduced antibody (~ 300 μL) in a 5:1 ratio and incubated at room temperature for 1.5 hours. For the second method, antibody-SMCC (~ 300 μL), was incubated with a 10-fold excess of BirA at room temperature for 2–6 hours. In both cases, the reactions were quenched with 100 μM β-mercaptoethanol for 30 minutes at room temperature, dialyzed against PBS, pH 7.2, and analyzed by Western blotting as described below.

Western blot analysis of antibody-BirA and antibody-AP conjugates

Conjugate preparations (15–30 μL) were combined with SDS-PAGE loading buffer containing 2-mercaptoethanol, boiled, run on a 10% SDS-PAGE gel, and transferred to nitrocellulose membrane. Membranes were blocked with TBS + 0.05% Tween 20 (TBS-T) and 3% BSA for 1 hour. For anti-goat IgG blotting, the membrane was first incubated with 20 μg/mL donkey anti-goat IgG horseradish peroxidase (Rockland Immunochemicals) in TBS-T + 3% BSA for 45 minutes, then washed three times. For anti-His_6 blotting, the membrane was first incubated with 10 μg/mL mouse anti-His_6 antibody (Calbiochem) in TBS-T + 3% BSA for 45 minutes, then washed three times. The membrane was then incubated with anti-mouse horseradish peroxidase conjugate (Bio-Rad) in TBS-T + 3% BSA at 1:2000 dilution for 45 minutes, followed by three rounds of 5 minute washes with TBS-T. Blots were developed with Supersignal West Pico or Femto substrate (Pierce), and images were taken on a ChemiImager 5500 instrument (Alpha Innotech).

Purification of antibody-biotin ligase conjugates by gel filtration

Biotin ligase-conjugates were purified from excess un-reacted BirA by gel filtration on a Superdex^{TM} 200 resin (GE Healthcare). We typically packed 6 mL of resin for a 250-275 μL of conjugation reaction, and the column was pre-equilibrated with 10 column volumes of PBS, pH 7.2 before loading the sample. The conjugate preparations were centrifuged (13,200 g, 3 minutes, room temperature) and loaded onto the column by gravity. The proteins were eluted from the column in 60×100 μL fractions in PBS, pH 7.2. Elution fractions were collected and analyzed by Western blotting as previously described.

Binding activity of antibody-acceptor peptide conjugates in live cells

Testing of AP conjugates to anti-mouse IgG was performed on HeLa cells using mouse anti-EGFR (clone LA22, Upstate, 1:25 dilution in DPBS + 3% BSA) as the primary antibody. AP-anti-rabbit conjugates were tested on HeLa cells stably expressing AP(ala)-CFP-TM[6] using rabbit polyclonal anti-HA (Rockland Immunochemicals, 1:50 dilution in DPBS + 3% BSA) as the primary antibody. In both cases, HeLa cells were incubated with primary antibody for 15 minutes at room temperature, followed by 15 minutes at room temperature with anti-mouse or anti-rabbit (Rockland Immunochemicals) conjugated to AP. Live cells were then rinsed with DPBS, and labeling was initiated by addition of 1 μM BirA, 10 μM biotin, and 500 μM ATP in DPBS. The biotinylation reaction was allowed to proceed for 30 minutes at room temperature. Biotin detection was performed by incubating the cells with 50 μg/mL streptavidin-Alexa Fluor 568 conjugate (Invitrogen) in DPBS + 3% BSA for 12 minutes at room temperature. After washing, the cells were imaged in DPBS with a Zeiss Axiovert 200M inverted epifluorescence microscope using a 40x oil-immersion lens. Alexa Fluor 568 (560/20 excitation, 585 dichroic, 605/30 emission), and differential interference contrast (DIC) images (630/10 emission) were collected and analyzed using OpenLab software (Improvision).

Binding activity of antibody-biotin ligase conjugates in live cells

Testing of biotin ligase conjugates to anti-mouse IgG was performed on HeLa cells using mouse anti-EGFR (clone LA22, Upstate, 1:25 dilution in DPBS-B) as the primary antibody. BirA-anti-rabbit conjugates were tested on HeLa cells stably expressing AP(ala)-CFP-TM[6] using rabbit polyclonal anti-HA (Rockland Immunochemicals, 1:50 dilution in DPBS-B) as the primary antibody. In both cases, HeLa cells were incubated with primary antibody, followed by anti-mouse or anti-rabbit antibodies (Rockland Immunochemicals) conjugated to BirA. Detection of anti-rabbit-BirA was performed by incubation with mouse monoclonal anti-His_6 antibody (Calbiochem, 1:10 dilution in DPBS-B), followed by anti-mouse antibody conjugated to Alexa Fluor 568 (Invitrogen, 6.7 μg/mL in DPBS-B). Detection of anti-mouse-BirA was performed by incubation with rabbit anti-BirA serum (21^{st} Century Biochemicals, 1:20,000 dilution in DPBS-B) followed by anti-rabbit antibody conjugated to Alexa Fluor 488 (Invitrogen, 6.7 μg/mL in DPBS-B). All antibody incubations steps were performed for 15 minutes at room temperature. Cells were rinsed and imaged as described above. The Alexa Fluor 488 filter set was 495/20 excitation, 515 dichroic, 530/30 emission.

Assay for proximity biotinylation of antibody-acceptor peptide conjugates by antibody-biotin ligase conjugates in live cells

HeLa cells were rinsed with DPBS prior to incubation with primary antibodies for 15 minutes at room temperature. Thereafter, cells were rinsed three times with DPBS and incubated for 15 minutes at room temperature with secondary antibodies conjugated to either AP or BirA. Subsequent labeling was performed for 30 minutes at room temperature in the presence of 10 μM biotin and 500 μM ATP, and the biotin was detected with 50 μg/mL streptavidin-Alexa Fluor 568 conjugate (Invitrogen) in DPBS-B for 12 minutes at room temperature. Cells were rinsed and imaged as described above. Two different tests were performed. In the first one, anti-rabbit-BirA biotinylation of anti-mouse-AP was tested on HeLa cells stably expressing the AP(ala)-CFP-TM plasmid[6] and using rabbit polyclonal anti-HA (Rockland Immunochemicals) and mouse anti-c-myc clone 9E10 (Calbiochem) as the primary antibodies. In the second assay, anti-mouse-

BirA biotinylation of anti-mouse-AP was tested on HeLa cells by staining endogenous EGFR with mouse anti-EGFR clone LA22 primary antibody (Upstate).

Fluorescent immunostaining of the epidermal growth factor receptor and some of its effector proteins

NR6, B82K+ or A431 cells were fixed with 3.7% paraformaldehyde (PFA) for 5 minutes at 4 °C, followed by 10 minutes at room temperature. Excess PFA was washed three times, 5 minutes each wash, with PBS and quenched with 100 mM glycine, pH 7.2 for 10 minutes at room temperature. After an additional 5-minute wash in PBS, cells were permeabilized by incubation with 100% methanol for 6 minutes at –20 °C. Finally, cells were rinsed twice for 10 minutes with PBS and blocked with 5% milk in PBS-Mg for 1 hour – overnight at room temperature or 4 °C with rocking. When appropriate, cells were serum-starved for 12–24 hours, and then EGF-stimulated by incubation with 100 ng/mL epidermal growth factor (EGF) for 10 minutes at 37 °C before fixation. Antigen detection was performed by incubating cells with 5 μg/mL of the appropriate primary antibody for 1 hour – overnight at room temperature in a humidified chamber. After three 10-minute washes with PBS-Mg + 0.1% Tween20, cells were incubated for 1 hour with 6.7 μg/mL anti-rabbit or anti-mouse antibody conjugated to either Alexa Fluor 488 or Alexa Fluor 568 (Invitrogen). After washing, the cells were imaged in PBS-Mg with a Zeiss Axiovert 200M inverted epifluorescence microscope using a 40x oil-immersion lens. Alexa Fluor 568 (560/20 excitation, 585 dichroic, 605/30 emission), Alexa Fluor 488 (495/20 excitation, 515 dichroic, 530/30 emission), and differential interference contrast (DIC) images (630/10 emission) were collected and analyzed using OpenLab software (Improvision). The following primary antibodies were used: rabbit polyclonal α-EGFR (Upstate, Cat# 06-847), mouse monoclonal α-EGFR Ab-12 (Lab Vision, Cat# MS-400-P1ABX, a cocktail with clones 111.6 + 199.12 + F4 + 11E8), mouse α-EGFR clone 199.12 (Lab vision, Cat# MS-396-P1ABX), mouse α-EGFR clone LA22 (Upstate, Cat# 05-104), rabbit polyclonal α-Shc (BD Biosciences, Cat# 610081), mouse α-Grb2 clone E-1 (Santa Cruz Biotechnology, Cat# sc-17813), α-Src clone GD11 (Upstate, Cat# 05-184), α-Crk II clone 18 (Santa Cruz Biotechnology, Cat# sc-289),

rabbit polyclonal α-Csk (Upstate, Cat# 06-566), rabbit polyclonal α-Nck (Chemicon, Cat# AB3167), and mouse α-phosphotyrosine clone 4G10 (Upstate, Cat# 05-321).

Assay for proximity biotinylation of antibody-acceptor peptide conjugates by antibody-biotin ligase conjugates in fixed cells

NR6 or A431 cells were stimulated, fixed, and blocked as described above for antigen immunostaining. Thereafter, blocking of endogenous biotin was performed by incubation with 50 μg/mL of avidin for 1 hour at room temperature, followed by blocking of the extra avidin sites by incubation with 100 μM biotin. Antigen detection was performed by incubating cells with 5 μg/mL of the appropriate primary antibody, followed by the corresponding secondary anti-mouse or anti-rabbit IgGs conjugated to biotin ligase and the acceptor peptide. After three 10-minute washes, proximity biotinylation was initiated by addition of 10 μM biotin and 500 μM ATP in DPBS for 1 hour at room temperature. Biotin detection was performed by incubating the cells with 5 μg/mL streptavidin-Alexa Fluor 568 conjugate (Invitrogen) in DPBS + 5% Milk + 0.03% Tween20 for 1 hour at 4 °C followed by three 10-minute washes with DPBS + 0.1% Tween20. After washing, cells were imaged as described above.

References

1. Miyashita,T. Confocal microscopy for intracellular co-localization of proteins. *Methods Mol. Biol.* **261**, 399-410 (2004).

2. Hell,S.W., Dyba,M. & Jakobs,S. Concepts for nanoscale resolution in fluorescence microscopy. *Curr. Opin. Neurobiol.* **14**, 599-609 (2004).

3. Kenworthy,A.K. Imaging protein-protein interactions using fluorescence resonance energy transfer microscopy. *Methods* **24**, 289-296 (2001).

4. Beckett,D., Kovaleva,E. & Schatz,P.J. A minimal peptide substrate in biotin holoenzyme synthetase-catalyzed biotinylation. *Protein Sci.* **8**, 921-929 (1999).

5. Diamandis,E.P. & Christopoulos,T.K. The biotin-(strept)avidin system: principles and applications in biotechnology. *Clin. Chem.* **37**, 625-636 (1991).

6. Chen,I., Howarth,M., Lin,W. & Ting,A.Y. Site-specific labeling of cell surface proteins with biophysical probes using biotin ligase. *Nat. Methods* **2**, 99-104 (2005).

7. Howarth,M., Takao,K., Hayashi,Y. & Ting,A.Y. Targeting quantum dots to surface proteins in living cells with biotin ligase. *Proc. Natl. Acad. Sci. U. S. A* **102**, 7583-7588 (2005).

8. de Boer,E. *et al.* Efficient biotinylation and single-step purification of tagged transcription factors in mammalian cells and transgenic mice. *Proc. Natl. Acad. Sci. U. S. A* **100**, 7480-7485 (2003).

9. Nezlin,R.S. The immunoglobulins: structure and function. Academic Press, (1998).

10. Hermanson,G.T. Bioconjugate Techniques. Academic Press, (1996).

11. Peluso,P. *et al.* Optimizing antibody immobilization strategies for the construction of protein microarrays. *Anal. Biochem.* **312**, 113-124 (2003).

12. Weaver,L.H., Kwon,K., Beckett,D. & Matthews,B.W. Corepressor-induced organization and assembly of the biotin repressor: a model for allosteric activation of a transcriptional regulator. *Proc. Natl. Acad. Sci. U. S. A* **98**, 6045-6050 (2001).

13. Eisenstein,E. & Beckett,D. Dimerization of the Escherichia coli biotin repressor: corepressor function in protein assembly. *Biochemistry* **38**, 13077-13084 (1999).

14. Chamow,S.M. *et al.* Conjugation of soluble CD4 without loss of biological activity via a novel carbohydrate-directed cross-linking reagent. *J. Biol. Chem.* **267**, 15916-15922 (1992).

15. Bieniarz,C., Husain,M., Barnes,G., King,C.A. & Welch,C.J. Extended length heterobifunctional coupling agents for protein conjugations. *Bioconjug. Chem.* **7**, 88-95 (1996).

16. Hashida,S., Imagawa,M., Inoue,S., Ruan,K.H. & Ishikawa,E. More useful maleimide compounds for the conjugation of Fab' to horseradish peroxidase through thiol groups in the hinge. *J. Appl. Biochem.* **6**, 56-63 (1984).

17. Xu,Y., Nenortas,E. & Beckett,D. Evidence for distinct ligand-bound conformational states of the multifunctional Escherichia coli repressor of biotin biosynthesis. *Biochemistry* **34**, 16624-16631 (1995).

18. Brock,R., Hamelers,I.H. & Jovin,T.M. Comparison of fixation protocols for adherent cultured cells applied to a GFP fusion protein of the epidermal growth factor receptor. *Cytometry* **35**, 353-362 (1999).

19. Chapman-Smith,A. & Cronan,J.E., Jr. Molecular biology of biotin attachment to proteins. *J. Nutr.* **129**, 477S-484S (1999).

20. Hollinshead,M., Sanderson,J. & Vaux,D.J. Anti-biotin antibodies offer superior organelle-specific labeling of mitochondria over avidin or streptavidin. *J. Histochem. Cytochem.* **45**, 1053-1057 (1997).

21. Livnah,O., Bayer,E.A., Wilchek,M. & Sussman,J.L. Three-dimensional structures of avidin and the avidin-biotin complex. *Proc. Natl. Acad. Sci. U. S. A* **90**, 5076-5080 (1993).

22. Yarden,Y. & Sliwkowski,M.X. Untangling the ErbB signalling network. *Nat. Rev. Mol. Cell Biol.* **2**, 127-137 (2001).

23. Schlessinger,J. Ligand-induced, receptor-mediated dimerization and activation of EGF receptor. *Cell* **110**, 669-672 (2002).

24. Ullrich,A. & Schlessinger,J. Signal transduction by receptors with tyrosine kinase activity. *Cell* **61**, 203-212 (1990).

25. Oksvold,M.P., Skarpen,E., Lindeman,B., Roos,N. & Huitfeldt,H.S. Immunocytochemical localization of Shc and activated EGF receptor in early endosomes after EGF stimulation of HeLa cells. *J. Histochem. Cytochem.* **48**, 21-33 (2000).

26. Lotti,L.V. *et al.* Sch proteins are localized on endoplasmic reticulum membranes and are redistributed after tyrosine kinase receptor activation. *Mol. Cell Biol.* **16**, 1946-1954 (1996).

27. Sakaguchi,K., Okabayashi,Y. & Kasuga,M. Shc mediates ligand-induced internalization of epidermal growth factor receptors. *Biochem. Biophys. Res. Commun.* **282**, 1154-1160 (2001).

28. Yang,S. *et al.* Mapping ErbB receptors on breast cancer cell membranes during signal transduction. *J. Cell Sci.* **120**, 2763-2773 (2007).

29. Sato,K. *et al.* Adaptor protein Shc undergoes translocation and mediates up-regulation of the tyrosine kinase c-Src in EGF-stimulated A431 cells. *Genes Cells* **5**, 749-764 (2000).

30. Suzuki,K. & Takahashi,K. Actin filament assembly and actin-myosin contractility are necessary for anchorage- and EGF-dependent activation of phospholipase Cgamma. *J. Cell Physiol* **189**, 64-71 (2001).

31. Langgut,W., Reisser,T., Kersten,H. & Nishimura,S. Modulation of epidermal growth factor receptor activity and related responses by the 7-deazaguanine derivative, queuine. *Oncogene* **8**, 3141-3147 (1993).

32. Santon,J.B. *et al.* Effects of epidermal growth factor receptor concentration on tumorigenicity of A431 cells in nude mice. *Cancer Res.* **46**, 4701-4705 (1986).

33. Pruss,R.M. & Herschman,H.R. Variants of 3T3 cells lacking mitogenic response to epidermal growth factor. *Proc. Natl. Acad. Sci. U. S. A* **74**, 3918-3921 (1977).

34. Reddy,C.C., Wells,A. & Lauffenburger,D.A. Proliferative response of fibroblasts expressing internalization-deficient epidermal growth factor (EGF) receptors is altered via differential EGF depletion effect. *Biotechnol. Prog.* **10**, 377-384 (1994).

35. Cull,M.G. & Schatz,P.J. Biotinylation of proteins in vivo and in vitro using small peptide tags. *Methods Enzymol.* **326**, 430-440 (2000).

36. Soderberg,O. *et al.* Direct observation of individual endogenous protein complexes in situ by proximity ligation. *Nat. Methods* **3**, 995-1000 (2006).

37. Gustafsdottir,S.M. *et al.* Proximity ligation assays for sensitive and specific protein analyses. *Anal. Biochem.* **345**, 2-9 (2005).

38. Jarvius,M. *et al.* In situ detection of phosphorylated platelet-derived growth factor receptor beta using a generalized proximity ligation method. *Mol. Cell Proteomics.* **6**, 1500-1509 (2007).

Chapter 4 : Proximity biotinylation for the detection of protein-protein interactions in live cells

The majority of the work discussed in this chapter is under review for publication at the *J. Am. Chem. Soc.* This project was initiated by Scott Chen, a graduate student in our lab. The experiments described in the first section of the results (Engineering BirA/AP for protein-protein interaction detection) as well as the determination of the rapamycin dose-response curve were performed by Scott. Yi Zhen assisted with the *in vitro* kinetic characterization of the BirA/AP(-3) pair.

Introduction

In this chapter, we describe the development of an alternative approach to protein-protein interaction (PPI) detection in cells, which offers a combination of spatial and temporal resolution with a low rate of false positives. In comparison to the methods described in Chapters 2 and 3, live-cell proximity biotinylation merges the advantages of temporal resolution (like the ribozyme-based methodology described in Chapter 2) and subcellular localization (like the enhanced immunofluorescence method described in Chapter 2); albeit at the expense of studying recombinant (not endogenous) proteins.

The method is similar to the protein complementation assays (PCAs) described in the introductory Chapter 1. In PCAs, two halves of a reporter protein are separately fused to each interaction partner, and reporter activity or fluorescence is restored by PPI-induced protein recombination[1]. Instead of two halves of a reporter protein, we proposed to use an enzyme-substrate pair (Figure 4-1). As described in Chapter 3, *E. coli* biotin ligase (BirA) is a 35 kD enzyme that catalyzes the covalent ligation of biotin to a lysine sidechain of a 15-amino acid recognition sequence called the "acceptor peptide" (AP)[2]. A crucial feature of the BirA/AP pair is that it is highly specific and orthogonal. When expressed inside or on the surface of mammalian cells, BirA biotinylates only AP fusion proteins and none of the endogenous mammalian proteins[3,4]. Conversely, AP is not recognized by mammalian biotin ligase[5].

In our scheme, the two proteins of interest (A and B) are fused to BirA and AP, respectively (Figure 4-1). Upon addition of biotin to cells, BirA will ligate biotin to the AP only if A and B are in close proximity. Subsequently, biotinylated AP can be detected with extremely high sensitivity and specificity on fixed cells by streptavidin staining[6], in cell lysate by streptavidin blotting[4,5], or on the surface of living cells by labeling with streptavidin-fluorophore or streptavidin-quantum dot conjugates[4,7-9]. The spatial resolution of this approach arises from the ability to image AP biotinylation through the use of streptavidin[4,7-9]. The temporal resolution comes from the rapid kinetics of enzymatic biotinylation (k_{cat} 12 min^{-1})[2], as well as the speed with which biotin can be delivered to the cytosol of mammalian cells, owing to the presence of biotin transporters at the plasma membrane[10].

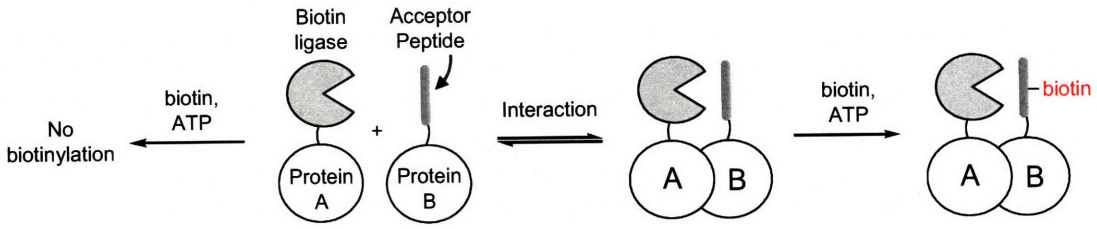


Figure 4-1: Proposed scheme for the detection of protein-protein interaction in living cells by proximity biotinylation. Each target protein (A and B) is fused to the *E. coli* biotin ligase enzyme (BirA) or its 15-amino acid acceptor peptide (AP). Interaction between proteins A and B brings BirA and AP into close proximity, resulting in the site-specific biotinylation of the acceptor peptide by biotin ligase.

Results

Engineering BirA/AP for protein-protein interaction detection

We first tested our methodology on a well-characterized interaction. For this proof of principle, we chose to study the interaction between the 12 kD FK506 binding protein (FKBP12) and the FKBP12-rapamycin binding domain of the mTor protein (FRB), whose interaction can be induced by addition of the small-molecule rapamycin[11]. There were several reasons to select this system. First, it is a high-affinity interaction that can be modulated by addition of a small molecule. On one hand, binding of rapamycin to FKBP12 has a dissociation constant of 0.2 nM[12]. On the other hand, binding of the FKBP12-rapamycin complex to FRB displays a K_d of 2–12 nM[11,13], whereas no binding has been observed *in vitro* or *in vivo* between FKBP12 and FRB in the absence of rapamycin[11]. This means that rapamycin can effectively act as an on-off switch of the interaction, which eases the characterization of the PPI detection method in terms of its rate of false positives and negatives. Second, the crystal structure of the FRB-rapamycin-FKBP12 ternary complex had been previously solved, which showed that the C-terminal ends of the two proteins are within ~ 18 Å of each other[14] (Figure 4-2). The crystal structure thus served as a guide to choose the optimal enzyme and substrate fusion sites, which bring the pair into closest proximity upon PPI. Third, the FKBP12-FRB interaction

had been previously used as a proof of principle by other PCA methods[15-20], which allowed us to compare features among different methodologies.

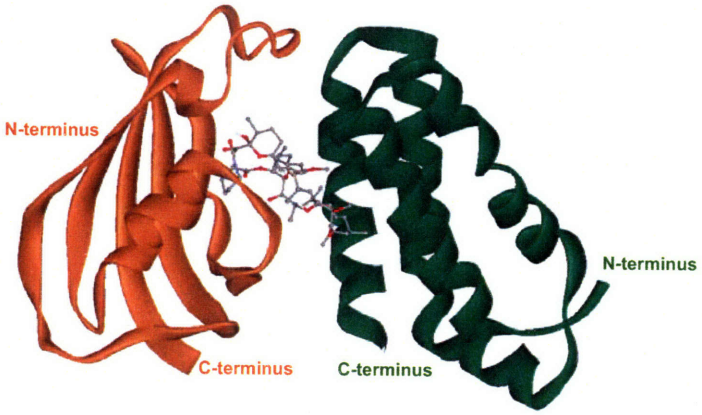


Figure 4-2: Crystal structure of the FKBP12-rapamycin-FRB ternary complex. The 12kD FK506 binding protein (FKBP12) has been colored orange and the FKBP12-rapamycin binding domain of the mTor protein (FRB) has been colored green. The N- and C-termina of each protein are labeled. The small molecule rapamycin, lying at the core of the complex, is represented in a ball and stick format. Carbon atoms are colored grey, oxygen atoms are colored red, and nitrogen atoms are colored blue. PDB source: 1FAP.

Guided by the crystal structure of the FKBP12-rapamycin-FRB complex, we thus created C-terminal fusions of FKBP12 and FRB to BirA and AP. In order to provide the enzyme and substrate pair with some flexibility to search for a productive conformation, we also introduced a 10-amino acid Gly-Ser intervening linker (Figure 4-5A; refer to the Experimental section for a complete description of the constructs). The fusion proteins were expressed in *E. coli*, purified, and tested *in vitro*. We incubated 1 μM of each fusion protein with 50 μM biotin in the presence or absence of 1 μM rapamycin, and we ran the reaction products on an SDS-PAGE, followed by detection of ligated biotin via streptavidin blotting. We observed stronger biotinylation of FKBP12-AP by FRB-BirA in the presence than in the absence of rapamycin (Figure 4-3A). Similar results were obtained when FRB was fused to AP and FKBP12 to BirA. Quantification of the streptavidin blot in Figure 4-3A yielded a signal/noise ratio in the presence/absence of rapamycin of ~ 5. Although this dynamic range seemed modest, we decided to next test the system in live cells, where the biotinylation levels could be dramatically different, both in the presence and absence of rapamycin. It must also be noted that incubation at

higher protein concentrations and longer biotinylation times both resulted in an increase in background, which significantly reduced the dynamic range of the *in vitro* assay.

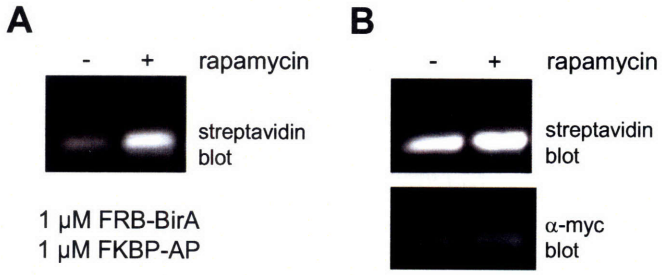


Figure 4-3: Proximity biotinylation *in vitro* and *in vivo* of full-length AP by FRB-BirA. (**A**) *In vitro* reporter characterization. Purified FRB-BirA and FKBP12-AP (original full-length AP) proteins were combined at 1 μM final concentration and incubated with 50 μM biotin in the presence or absence of 1 μM rapamycin. The extent of biotinylation of FKBP12-AP was analyzed by streptavidin blot. (**B**) *In vivo* reporter characterization. HEK cells co-transfected with FRB-BirA and FKBP12-AP were either treated with 100 nM rapamycin for 1 hour or left untreated. Biotin was then added to all cells for 1 minute before lysis and analysis by streptavidin blotting. The lower blot (α-myc) stains the myc tag in the FKBP12-AP construct.

Figure 4-4 depicts our protocol for testing the system in live cells. We transfected human embryonic kidney (HEK) cells with the two fusion proteins and, 48 hours after transfection, we treated the cells with rapamycin for 1 hour, followed by 50 μM biotin for 1 minute. The cells were lysed, and ligated biotin was detected by streptavidin blotting.

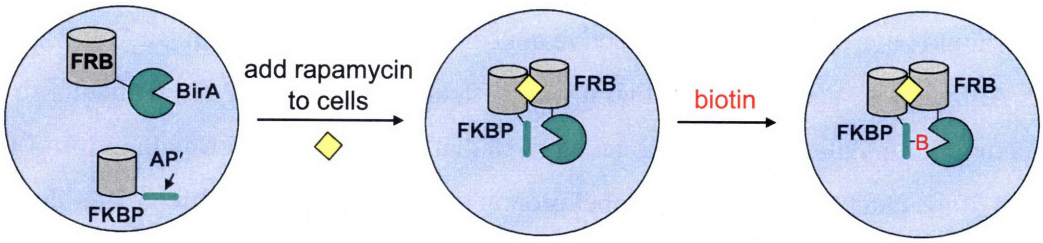


Figure 4-4: Live cell detection of rapamycin-mediated FKBP12-FRB interaction. Cells are first transfected with both fusion constructs; in this case, FRB-BirA and FKBP12-AP′. After transfection, 100 nM rapamycin is added to cells for 1–3 hours to promote protein-protein interaction. Thereafter, 50 μM biotin is added to the cell media. Biotin enters the cells very rapidly via cell-surface biotin transporters. Once inside, biotin is recognized by FRB-BirA, which also uses endogenous ATP to biotinylate FKBP12-AP′. Ligated biotin can then be detected by Western blotting of cell lysates or by immunostaining of fixed cells.

We observed little difference in biotinylation intensity with and without rapamycin (Figure 4-3B). Because no interaction has ever been detected between FRB and FKBP12 in the absence of rapamycin, it was unlikely that background biotinylation

was being caused by non-rapamycin dependent FRB-FKBP12 interactions. We hypothesized then that the protein expression levels in HEK were much greater than the protein concentrations used for our *in vitro* studies, which resulted in BirA/AP interaction even in the absence of interaction between FRB and FKBP12, ultimately giving rise to increased background biotinylation. The reported K_M for the interaction between BirA and AP is 25 μM (Ref. 2), which is close to the recombinant protein concentrations that can be expected upon transient transfection in HEK cells. Hence, our first approach to decrease background biotinylation was to reduce intracellular protein concentrations. We achieved this by reducing the amount of DNA used for transfection. Although reduced DNA resulted in decreased protein concentrations as observed by α-myc (FKBP12-AP) and α-HA (FRB-BirA) blotting, the signal/background ratio remained almost unchanged (data not shown). Other variations to the experimental protocol, such as depleting the cell growth media from biotin after transfection or decreasing biotin concentration during the labeling, all failed to increase the +/– rapamycin signal ratio in cells to greater than ~ 1.3. We thus decided to shift our attention toward a more fundamental change in the system, that of the interaction affinity between BirA and AP.

We reasoned that reduction of BirA affinity for AP (i.e., increase in K_M) could preferentially decrease the extent of biotinylation of non-complexed AP compared to complexed AP. This was based on the fact that complexed AP, within the context of a PPI, should have a much higher effective concentration with respect to BirA than free AP. Although we acknowledged that it was difficult to engineer AP K_M without also affecting biotinylation k_{cat}, we also recognized that we had some leeway in terms of the speed of the catalytic reaction. To label protein complexes with half-life greater than 1 minute, we required only a biotinylation k_{cat} of 1 min^{-1} or greater; which matched well with our typical time of biotin delivery to living cells of ~ 1 minute.

To engineer the affinity of the BirA/AP interaction, we truncated the 15-amino acid sequence of the AP at the C-terminus, the N-terminus, or both (Figure 4-5A). We hoped that these truncations would decrease the interaction surface area with the BirA active site, potentially leading to a higher K_M. Although it was hard to predict how these mutations would affect the rate of catalysis, we expected that at least some activity would be retained, since the residues around the catalytic lysine were not perturbed. We

fused the truncated AP variants to FKBP12 and we expressed them in HEK cells, together with FRB-BirA. We observed a wide range of biotinylation intensities in the presence and absence of rapamycin (Figure 4-5B). We were delighted to see that the signal/background ratio was increased for all the truncated variants compared to that of the full-length AP. For all the AP truncation mutants, the background level of biotinylation in the absence of rapamycin did decrease. However, many times the +rapamycin signal also decreased. The best construct was the AP(-3) peptide, with three residues deleted from the C-terminal end of AP, which displayed the lowest –rapamycin signal but a +rapamycin signal comparable in intensity to that of the original full-length AP (signal/background ratio ~ 8.5). We selected this new peptide for further characterization.

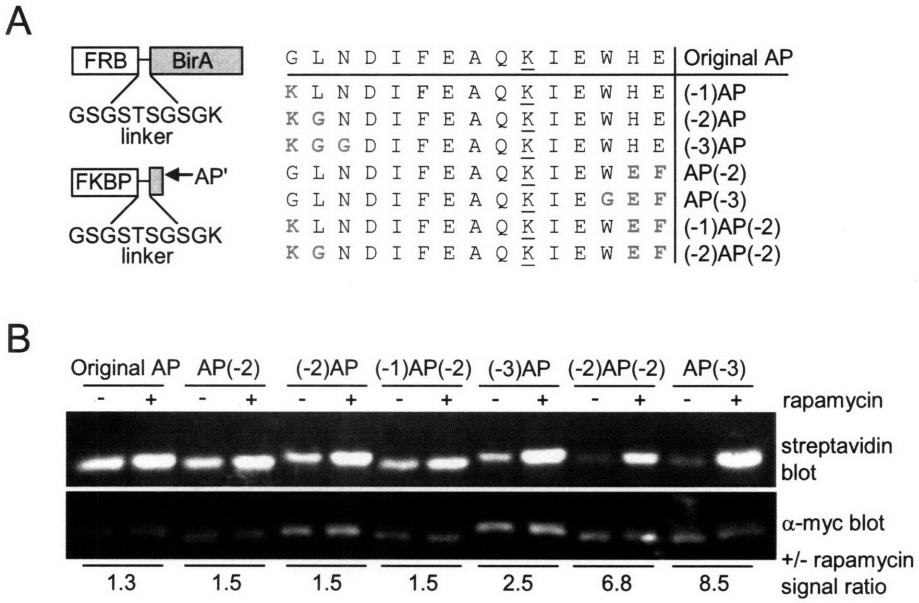


Figure 4-5: Engineering the BirA/AP pair. (**A**) Left: domain structures of constructs used to test the methodology *in vitro* and in living cells. Right: FKBP12 was fused to the original AP sequence as well as to 7 truncated variants of the AP. The lysine biotinylation site is underlined. Placeholder amino acids inserted at truncation sites are colored green. (**B**) Comparison of AP mutants in a live cell assay. HEK cells co-transfected with FRB-BirA and one of the FKBP12 fusions to a truncated AP (FKBP12-AP′) were either treated with 100 nM rapamycin for 1 hour or left untreated. Biotin was then added to all cells for 1 minute before lysis and analysis by streptavidin blotting. The lower blot (α-myc) stains the myc tag in each of the FKBP12-AP′ constructs. The +/– rapamycin ratio was calculated from the streptavidin blot signals.

Characterization of the new BirA/AP(-3) pair

We first performed a series of optimizations to maximize the +/- rapamycin signal intensity for the FRB-BirA/FKBP12-AP(-3) pair. It had been previously reported that the effect of rapamycin when added to mammalian cells reaches a maximum after ~ 3 hours[16]. Increasing the rapamycin incubation time from 1 to 3 hours, in combination with reduced protein expression levels, resulted in an increase of the signal/background ratio to ~ 12.5 (Figure 4-6, lanes 1 and 2) from the original 8.5 (Figure 4-5B, last two lanes). Second, we used a mitochondrial activity test to determine the effect of biotin-starvation on cell health. We found that shortening the time of incubation of cells in biotin-free media from the original 24 hours to ~ 12 hours, resulted in an optimal balance between the requirements for minimal background biotinylation of AP(-3), and maintenance of cell health. With these new conditions, biotin-starved cells exhibited a 99% ± 5% of surviving cells when compared to control untransfected cells. Notably, transient transfection of the same cells resulted in higher toxicity (86% ± 4% of surviving cells) than that observed for biotin-starvation. Third, we performed several tests of labeling specificity in live HEK cells. Figure 4-6 shows that biotinylation of FKBP12-AP(-3) is both site-specific (K→A mutant, lane 4) and dependent on BirA activity (K183R mutant, lane 5). We also showed that omission of biotin resulted in no modification of FKBP12-AP(-3) (lane 3).

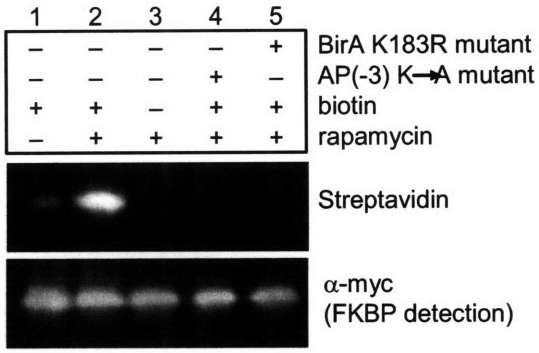


Figure 4-6: Analysis of detection specificity in living cells. HEK cells co-transfected with FRB-BirA and FKBP12-AP(-3) were treated with rapamycin and biotin, lysed, and analyzed by streptavidin blotting. Lanes 1 and 2 show the increase in FKBP12-AP(-3) biotinylation upon rapamycin addition. Lane 3 is a negative control where biotin was omitted. Lane 4 shows a negative control with FKBP12-AP(-3) replaced by an alanine mutant at the lysine biotinylation site. Lane 5 shows a negative control in which the FRB-BirA fusion was replaced with a

catalytically-inactive mutant, FRB-BirA (K183R). The α-myc blot demonstrates equal expression of FKBP12-AP(-3) in all lanes.

Next, we tested the fusion proteins *in vitro* and observed good signal to background ratios at two different protein concentrations, 100 nM and 1 μM (Figure 4-7). Notably, the –rapamycin background signal using 1 μM protein was much lower than that obtained using the original BirA/AP pair (Figure 4-3A), which resulted in an increase in the signal/background ratio *in vitro* from ~ 5 for the original AP to ~ 23 for AP(-3). The strong +rapamycin signal obtained at 100 nM protein demonstrates the high sensitivity of this reporter system at low protein concentrations.

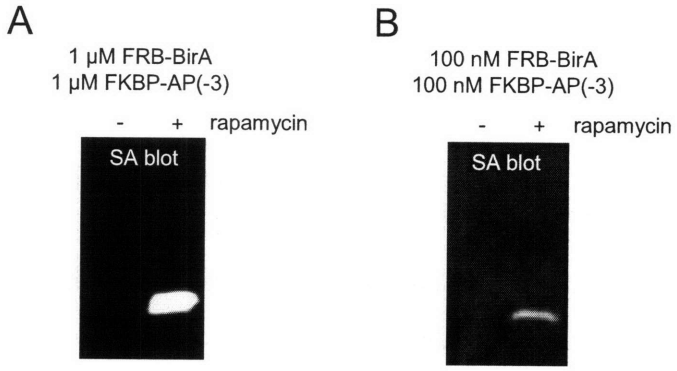


Figure 4-7: *In vitro* characterization of the BirA/AP(-3) pair. Purified FRB-BirA and FKBP12-AP(-3) proteins were combined at either 1 μM (A) or 100 nM (B). After incubation with 50 μM biotin for 1 minute (A) or 5 minutes (B) in the presence or absence of 1 μM rapamycin, the extent of biotinylation of FKBP12-AP(-3) was analyzed by streptavidin blot. The +/- rapamycin signal intensity ratios are ~ 23 (A) and ~ 17 (B).

We also performed kinetic measurements to characterize the BirA/AP(-3) pair. For this purpose, we fused AP(-3) to HP1, a chromodomain protein that we were able to express at very high levels in *E. coli*, and that had been previously used in our lab to determine the biotinylation kinetics of a different acceptor peptide by yeast biotin ligase[21]. We used a ^{3}H-biotin filter binding assay to estimate the Michaelis-Menten curve (Figure 4-8). We obtained a K_M of 345 ± 19 μM and a k_{cat} of 0.53 ± 0.01 min^{-1}. The new K_M was 14-fold greater than the original K_M of 25 μM for full-length AP[2]. This increase in K_M explained the lower background in the cellular experiments. Although the k_{cat} for ligation of biotin to AP(-3) was 23-fold slower than to AP (12 min^{-1}) (Ref. 2), the reaction should still be fast enough to robustly label protein complexes with half-life greater than ~ 1 minute in cells.

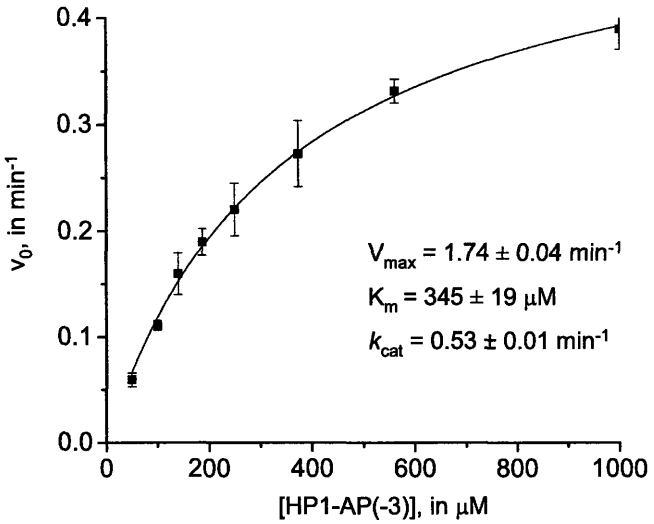


Figure 4-8: Kinetics of AP(-3) biotinylation by BirA. The Michaelis-Menten curve shows the initial rates of BirA biotinylation of an HP1-AP(-3) fusion protein, as a function of HP1-AP(-3) concentration. BirA concentration was 3.3 μM. Each data point represents the average of three independent experiments. Error bars, ± 1 s.d.

An ideal PPI detection method should also be capable of estimating interaction K_ds. To test if our method was capable of such quantification, we obtained a dose-response curve for rapamycin in living cells. HEK cells co-expressing FRB-BirA and FKBP12-AP(-3) were treated with various concentrations of rapamycin, lysed, and analyzed by streptavidin blotting. The dose-response curve in Figure 4-9 shows a saturable single-site binding curve with an apparent K_d of 10 nM, which matches the reported *in vitro* K_d for the FKBP12-rapamycin interaction with FRB of 12 nM[13]. Our results are also in accordance with those obtained using the split-luciferase (K_d = 1.5 ± 0.3)[16] and split-β-lactamase (K_d 5 nM)[15] methodologies.

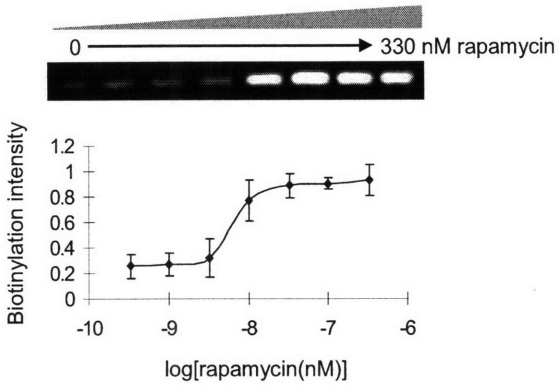


Figure 4-9: Dose-response curve of rapamycin in live cells. The sigmoidal curve shows the biotinylation response to increasing concentrations of rapamycin. HEK cells were co-transfected with FRB-BirA and FKBP12-AP(-3), incubated with various concentrations of rapamycin for 1 hour, then treated with biotin for 1 minute before lysis. The graph shows the normalized FKBP12-AP(-3) biotinylation signal intensities from three independent experiments. Error bars, ± 1 s.d.

Optimizing the BirA/AP(-3) pair for the detection of protein-protein interactions in mammalian cells

Although engineering the affinity between BirA and AP did indeed result in an increase of the signal/background ratio, it is clear from Figure 4-6 that certain non-PPI-mediated interaction between BirA and AP(-3) still remained when the constructs were expressed at high concentrations in mammalian cells. We thus wished to further decrease the rate of false positives (i.e., –rapamycin signal) by controlling the concentrations of the FRB-BirA and FKBP12-AP(-3) proteins.

In the previous section, we attempted to decrease protein concentration by decreasing the amount of DNA added during tranfection. In our experience, however, this approach has a very limited dynamic range, and it is only successful for moving from extremely high to high or medium protein concentrations. We then decided to transfer the FRB-BirA and FKBP12-AP(-3) constructs into a plasmid that expresses the protein under the regulation of a tetracycline operator. Specifically, we used the T-REx™ System developed by Invitrogen. Tetracycline regulation in the T-REx™ System is based on the binding of tetracycline to the Tet repressor, which releases the repressor and results in derepression of the promoter controlling expression of the gene of interest[22]. We cloned the constructs into the tetracycline-inducible plasmid pcDNA4-Tet and we transfected

HEK cells stably expressing the Tet repressor (T-REx^{TM}-293). We performed the proximity biotinylation assay as before, except that varying amounts of tetracycline were added to the cell media after transfection. Briefly, right after transfection, we incubated the cells in biotin-free media containing tetracycline for ~ 16–24 hours. Thereafter, we treated the cells with rapamycin for 3 hours, followed by 50 μM biotin for 1 minute. The cells were lysed, and ligated biotin was detected by streptavidin blotting. As expected, reducing protein concentrations resulted in a marked increased in the signal/background ratio, albeit to the expense of reducing the overall signal (Figure 4-10). In the case of the FRB-FKBP12 interaction, a tetracycline concentration of ~ 0.3 μg/mL provided an optinal balance between minimal false positives and a detectable signal. The optimal tetracycline concentration would need to be determined for every new protein-protein interaction under study.

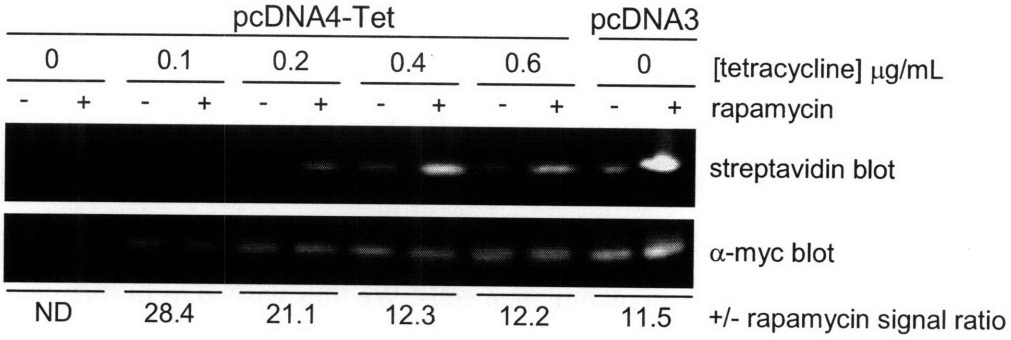


Figure 4-10: Titration of protein concentration in the proximity biotinylation detection of PPIs. T-REx^{TM}-293 cells were co-transfected with FRB-BirA and FKBP12-AP(-3), and protein expression was induced with varying amounts of tetracycline ranging from 0–0.6 μg/mL. Cells were then treated with rapamycin for 3 hours (+rapamycin) or left untreated (–rapamcyin). Biotin was then added to all cells for 1 minute before lysis and analysis by streptavidin blotting. The α-myc blot confirms that the expression levels of FKBP12-AP(-3) increased with increasing concentrations of tetracycline. The +/– rapamycin ratio was calculated from the streptavidin blot signals. The same experiment was also performed in parallel on regular HEK cells transfected with the original non-tetracycline inducible plasmid (pcDNA3). The results are shown for comparison (last two lanes).

Thus, the new BirA/AP(-3) pair was able to, under optimal conditions, result in a ~ 25-fold signal induction upon addition of rapamycin. As previously mentioned, other PPIs methods have used the same proof of principle interaction, which allowed us to compare our results. Only the split-β-galactosidase system was reported to result in a

higher fold of signal induction (~ 30-fold) upon addition of rapamycin to C2C12 cells[19]. However, the high signal induction afforded by the β-galactosidase system is a reflection of its enzymatic signal amplification, which, although useful for enhancing sensitivity, renders the system not amenable to quantification and can potentially increase the rate of false positives. In contrast, the two most popular PCAs, split-β-lactamase and split-luciferase, both result, under similar conditions, in lower signal induction folds (~ 10-fold for β-lactamase[20] and ~ 6-fold for luciferase[16]).

Finally, we used a mobility shift assay to quantify the extent of FKBP12-AP(-3) biotinylation under different tetracycline concentrations (Figure 4-11). T-REx™-293 or HEK cells were transfected and treated with rapamycin and biotin as before. After lysis, cell extracts were incubated with excess soluble streptavidin, which bound biotinylated FKBP12-AP(-3). Binding mixtures were then run in an SDS-PAGE without boiling to preserve the biotin-streptavidin binding, and both, unmodified and biotinylated protein, were detected by α-myc blotting. Figure 4-11 shows the mobility shift caused by streptavidin binding to biotinylated FKBP-AP(-3). It must be noted that, owing to the size difference, the efficiency of the transfer to nitrocellulose varies for the biotinylated and unmodified proteins, which explains the much lower α-myc signal for the streptavidin-bound protein. For this reason, we quantified the extent of biotinylation by comparing the α-myc signal for unmodified FKBP12-AP(-3) (lower bands) in the – and + streptavidin lanes, at each tetracycline concentration. The results are expressed as the percentage of the total FKBP12-AP(-3) that was biotinylated in the cells (Figure 4-11). The percentage biotinylation ranged from 6–8%. As expected, no significant differences were observed for the different tetracycline concentrations. These results are thus in agreement with the previous ones (Figure 4-10). Since the percentage of biotinylated protein remains relatively constant across tetracycline concentrations, the overall biotinylation signal increases solely as a function of the total protein concentration.

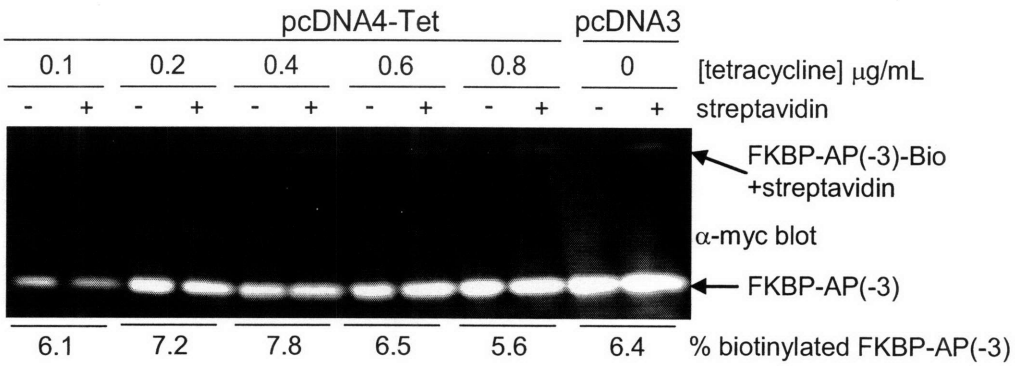


Figure 4-11: Quantification of the extent of biotinylation of FKBP12-AP(-3) at different cellular protein concentrations. T-REx-HEK cells were co-transfected with FRB-BirA and FKBP12-AP(-3), and protein expression was induced with varying amounts of tetracycline ranging from 0.1–0.8 μg/mL. Cells were then treated with rapamycin and biotin, and lysed. The lysates were incubated in the presence or absence of excess streptavidin (–/+ streptavidin) and analyzed by α-myc blotting. Biotinylated protein (FKBP12-AP(-3)-Bio) is separated from unmodified protein based on the mobility shift caused by streptavidin binding to biotin. The upper bands correspond to shifted, biotinylated protein, whereas the lower bands correspond to unmodified FKBP12-AP(-3). The extent of biotinylation was calculated from the –/+ streptavidin difference in the α-myc signal of unmodified FKBP12-AP(-3) (lower bands). The same experiment was also performed in parallel on regular HEK cells transfected with the original non-tetracycline inducible plasmid (pcDNA3). The results are shown for comparison (last two lanes).

Imaging the rapamycin-dependent interaction between FRB and FKBP12 in mammalian cells

One of the main features of the proximity biotinylation method is its ability to provide information on the subcellular localization of the PPI. This is the case because, in contrast to the diffusible small-molecule substrates of luciferase and β-lactamase, biotin remains covalently bound to one of the interacting partners. To demonstrate this feature, we once again performed detection of the FRB-FKBP12 interaction in live cells, but we then detected the ligated biotin by streptavidin staining of fixed cells instead of western blotting. Specifically, HEK cells co-expressing FRB-BirA and FKBP12-AP(-3) were incubated in biotin-free media for 12 hours, to minimize background both from AP(-3) and from the four endogenous protein substrates of mammalian biotin ligase[23]. The cells were then treated with 100 nM rapamycin for 3 hours, followed by 50 μM biotin for 1 minute, then fixed and stained with streptavidin conjugated to Alexa Fluor 568.

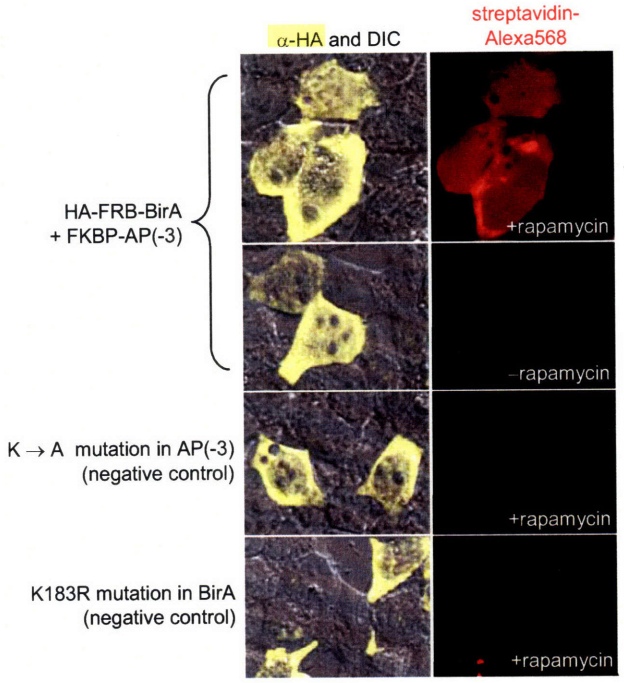


Figure 4-12: Imaging the FRB-FKBP12 interaction in the cytoplasm of HEK cells. HEK cells were co-transfected with FRB-BirA and FKBP12-AP(-3). After 3 hours incubation with 100 nM rapamycin, cells were labeled with biotin for 1 minute, then fixed and permeabilized with formaldehyde/methanol, and stained with streptavidin-Alexa Fluor 568. Anti-HA staining was also performed to detect expression of HA-FRB-BirA. Streptavidin labeling is shown in red to the right of merged DIC (differential interference contrast) and anti-HA (yellow) images. Controls are shown with rapamycin omitted, an alanine mutation in AP(-3), and a K183R mutation in BirA.

The top row of Figure 4-12 shows that transfected cells, as indicated by immunostaining of the HA epitope in the FRB-BirA protein, were biotinylated. In contrast, neighboring untransfected cells in the same field of view show no signal. Quantification of several samples gave a +/–rapamycin signal ratio of ~ 5.1. This 5-fold induction with rapamycin was similar to results obtained with split-luciferase technology under similar conditions[16]. Although streptavidin staining of transfected cells in the –rapamycin control was above background staining of untransfected cells, the signal on those cells was always significantly lower than the signal observed in the +rapamycin samples. This demonstrated the low rate of false positives that characterizes the proximity biotinylation assay. A negative control where the lysine site of modification in AP was mutated to an alanine showed no labeling, thus confirming the site-specificity of the labeling. We also performed an additional control by mutating Lysine 183 in BirA,

which had been previously suggested to be involved in catalysis[24]. The Lys183Arg mutation negative control also showed no labeling, which confirmed the dependence of the labeling on BirA activity.

To further illustrate the capability of proximity biotinylation to provide subcellular localization, we appended a nuclear localization signal (NLS) to FRB-BirA and FKBP12-AP(-3). We performed the imaging experiment as before, except that we used the tetracycline-inducible system, and treated the T-REx^{TM}-293 cells with 0.4 μg/mL tetracycline, concentration that we found optimal for the nuclear proteins. The results are shown in Figure 4-13.

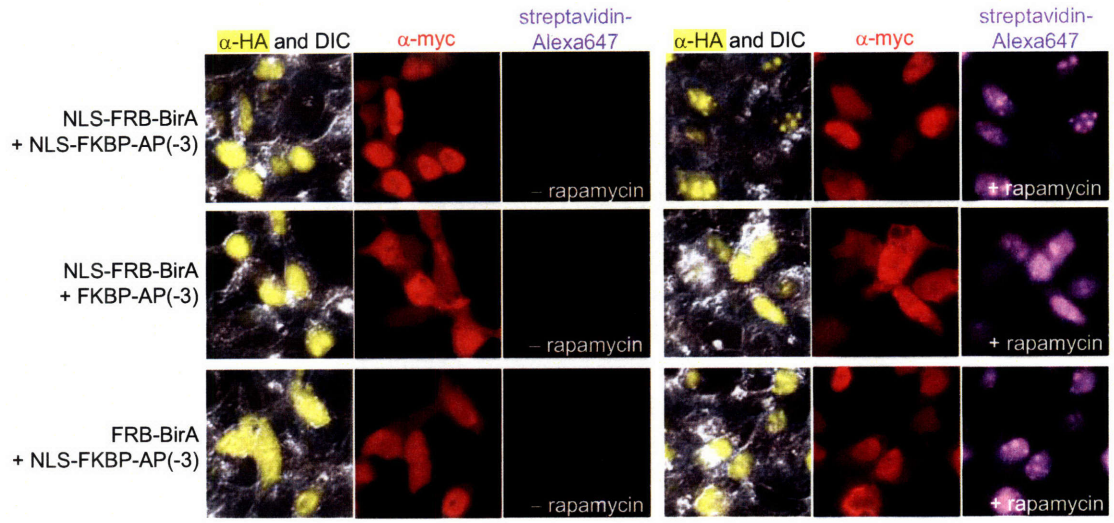


Figure 4-13: Imaging the FRB-FKBP12 interaction in the nucleus of HEK cells. T-Rex HEK cells were co-transfected with the indicated nuclear or cytosolic FRB-BirA and FKBP12-AP(-3) plasmids, and protein expression was induced with 0.4 μg/mL tetracycline for 16–24 hours. After 3 hours incubation with 100 nM rapamycin, cells were labeled with biotin for 1 minute, then fixed and permeabilized with formaldehyde/methanol, and stained with streptavidin-Alexa Fluor 568. Anti-HA and anti-myc stainings were also performed to detect expression of HA-FRB-BirA and Myc-FKBP12-AP(-3). Streptavidin labeling is shown in purple to the right of merged DIC (differential interference contrast) and anti-HA (yellow) images, and the anti-myc (red) images.

We observed clear nuclear staining when both, FRB-BirA and FKBP12-AP(-3) were targeted to the nucleus (Figure 4-13, upper panels). The streptavidin signal was highly concentrated in the nucleoli, where FRB-BirA was also in greater abundance. As expected, the streptavidin staining remained nuclear when the nuclear constructs were combined with the non-targeted ones (Figure 4-13, middle and lower panels). Notably,

FRB-BirA translocated to the nucleus when co-expressed with NLS-FKBP12-AP(-3) only in the presence, but not in the absence, of rapamycin treatment.

Conclusions

In this chapter, we have described a new methodology to detect PPIs *in vitro* and in living cells via an enzyme-substrate interaction rather than reporter protein reconstitution. We tuned the BirA/AP affinity to reduce background and eliminate false positives, while still allowing robust detection of relatively transient PPIs (half-life > 1 minute). We demonstrated that the methodology exhibits high specificity for the detection of PPIs in living mammalian cells, with a fold induction in the detected signal upon PPI of ~ 5–25, depending on whether imaging or Western blotting were used for readout. We also showed that careful control of protein concentration in cells (e.g., using a tetracycline-inducible system) resulted in a ~ 2.5-fold increase in the dynamic range of the assay. As seen for the FRB-FKBP12 system, the BirA/AP(-3) pair was also able to quantitatively predict interaction K_ds. Importantly, we showed that proximity biotinylation can detect the subcellular localization of the PPI under study, at least in the cytosol and the nucleus of mammalian cells.

Table 4-1 compares the features of proximity biotinylation to those of the most popular methods to detect cellular PPIs. Although the methodology is restricted to analysis of a single time-point, due to the requirement for cell fixation or lysis before streptavidin detection, it still bridges many of the positive features of existing PCAs, including good spatial resolution and rapid response time in mammalian cells. Biotin is readily taken up by all cell types by a combination of passive diffusion and active transport via biotin transporters[10]. In contrast, the split-β-lactamase technology is limited by the impermeability of its small-molecule substrate in many cell types[25]. An additional improvement over most PCAs is the reversible nature of the BirA/AP interaction, which minimizes perturbation to the PPI under study.

Last, as it was mentioned in the introductory Chapter 1, an important extension of any cellular PPI detection method is the discovery of new protein interaction partners

through library screens. Proximity biotinylation should be particularly well-suited for this application because the biotin label used for detection can also be a handle for protein purification using streptavidin-conjugated beads.

Table 4-1: Comparison of proximity biotinylation to FRET and protein fragment complementation assays. (1) Although, in general the temporal resolution of split-GFP is limited by the time required for GFP maturation, a recently developed split-GFP system was reported to mature within minutes of DNA hydridization[26]. (2) Most reported split-luciferase systems result in the irreversible re-constitution of luciferase activity; however, a split-luciferase system capable of quantifying dynamic interactions was recently reported[27].

Strategy	Rate of false positives	Rate of false negatives	Subcellular localization	Real-time Readout	Traps interaction	Living animals
Intermolecular FRET	Moderate	High	Yes	Yes	No	No
Split-GFP	High	Low	Yes	No (1)	Yes	No
Split-β-lactamase	High	Low	No	Yes	Yes	No
Split-luciferase	High	Low	No	Yes	Yes (2)	Yes
Proximity biotinylation	Low	Low	Yes	No	No	No

Experimental

Cloning and mutagenesis

Note: nucleotides encoding AP sequences are underlined, and nucleotides encoding epitope tags are italicized.

FRB-BirA-pET21b bacterial expression plasmid

The human FRB gene (a gift from Paul Clemons, Broad Institute) was PCR-amplified using the primers 5′-AAA AGG ATC CGT *ATC CGT ACG ACG TAC CAG ACT ACG* CAA TGT GGC ATG AAG GCC TG and 5′-TTT TGA ATT CTT TGC CCG AGC CCG AGG TCG AGC CCG AGC CCT TTG AGA TTC GTC GGA A to introduce a BamHI site and an HA tag (YPYDVPDYA) at the N-terminus, and a 10-amino acid linker (GSGSTSGSGK) and an EcoRI site at the C-terminus. The *E. coli* BirA gene was PCR-amplified using the primers: 5′-AAA AGA ATT CAT GAA GGA TAA CAC CGT G and 5′-TTT TAA GCT TAT TTT TCT GCA CTA CG, digested with EcoRI and HindIII, and pasted into an EcoRI/HindIII-digested pRSETb vector (Invitrogen) to create a BirA-pRSETB plasmid. The FRB PCR product was then digested with BamHI/EcoRI and ligated into BamHI/EcoRI-digested BirA-pRSETb plasmid, to give an FRB-BirA-pRSETb plasmid. This gene was then PCR-amplified with the primers: 5′-AAA AGG ATC CGT *ATC CGT ACG ACG TAC CAG ACT ACG* CAG AGA TGT GGC ATG AAG GCC TG and 5′-TTT TAA GCT TTT TTT CTG CAC TAC GCA GGG ATA TT and ligated into the BamHI and HindIII sites of pET21b (Novagen) to give FRB-BirA-pET21b.

FKBP12-AP-pET21b and FKBP12-AP(-3)-pET21b bacterial expression plasmids

PCR of the human FKBP12 gene (a gift from Paul Clemons, Broad Institute) was performed using the primers 5′-GAT AAG GAT CCG *GAA CAA AAA CTT ATT TCT GAA GAA GAT CTG* GGA GTG CCA GGT GGA AAC CAT CTC C and 5′-CAT CAT GAA TTC CTC GTG CCA CTC GAT CTT CTG GGC CTC GAA GAT ATC GTT CAG GCC TTT GCC CGA GCC CGA GGT CGA GCC CGA GCC TTC CAG TTT TAG AAG CTC CAC ATC G to introduce a myc tag (EQKLISEEDL) and a BamHI site at the N-terminus, and the AP sequence

(GLNDIFEAQKIEWHE), a 10-amino acid linker (GSGSTSGSGK), and an EcoRI site at the C-terminus. The PCR product was ligated into the BamHI and EcoRI sites of pRSETb (Invitrogen), to obtain FKBP12-AP-pRSETb.

Using this plasmid as a template, a second PCR was performed using the primers 5′-GAT AAG GAT CCG *GAA CAA AAA CTT ATT TCT GAA GAA GAT CTG* GGA GTG CCA GGT GGA AAC CAT CTC C and 5′-TTT TGA ATT CTC GTG CCA CTC GAT CTT CTG GGC. The PCR product was digested, and ligated into the BamHI and EcoRI sites of pET21b (Novagen) to obtain FKBP12-AP-pET21b.

FKBP12-AP(-3)-pET21b was created using FKBP12-AP-pET21b as a PCR template and the primers 5′-GAT AAG GAT CCG *GAA CAA AAA CTT ATT TCT GAA GAA GAT CTG* GGA GTG CAG GTG GAA ACC ATC TCC and 5′-TTT TGA ATT CTT AGA ACT CCC CCT CGA TCT TCT GGG CCT CG. The PCR product was digested, and ligated into the BamHI and EcoRI sites of pET21b (Novagen).

FRB-BirA-pcDNA3 mammalian expression plasmid

To create the FRB-BirA-pcDNA3 construct, the EcoRI site in FRB-BirA-pRSETB was first abolished by QuikChange using the forward primer 5′-GGG CTC GGG CAA AGA GTT CAT GAA GGA TAA CAC and its reverse complement. The FRB-BirA gene was then amplified by PCR using the primers 5′-AAA AGG ATC CG*T ATC CGT ACG ACG TAC CAG ACT ACG CAG* AGA TGT GGC ATG AAG GCC TG and 5′-TTT TGA ATT CTT TTT CTG CAC TAC GCA GGG ATA TT, followed by ligation into the BamHI and EcoRI sites of pcDNA3 (Invitrogen).

FKBP12-AP-pcDNA3 and FKBP12-AP(-3)-pcDNA3 mammalian expression plasmids

Both constructs were created by PCR of the corresponding pET21b templates using the primers 5′-GAT AAG GAT CCG *GAA CAA AAA CTT ATT TCT GAA GAA GAT CTG* GGA GTG CAG GTG GAA ACC ATC TCC and 5′-TTT TGA ATT CCC CCT CGA TCT TCT GGG CCT CG followed by ligation of the PCR product into the BamHI and EcoRI sites of pcDNA3 (Invitrogen).

HP1-AP(-3)-pET21a bacterial expression plasmid

The HP1 gene was PCR-amplified with the primers: 5′-TTT CGC GGA TCC GCA CCA CCA CCA CCA CCA CGA GGA GGA GTA CGC CGT G and 5′-TTT TGA ATT CAT TAG AAC TCC CCC TCG ATC TTC TGG GCC TCG to introduce a BamHI site and a His_6 tag at the N-terminus, and the AP(-3) tag and an EcoRI site at the C-terminus. The PCR product was ligated into the BamHI and EcoRI sites of pET21a (Novagen).

Site directed mutagenesis on AP and BirA

The alanine mutant of the FKBP12-AP(-3)-pcDNA3 plasmid was obtained by QuikChange using 5′-GAT ATC TTC GAG GCC CAG GCG ATC GAG GGG GAA TTC and its reverse complement. The K183 mutation on FRB-BirA-pcDNA3 was performed by QuikChange using 5′-CTG CAG GAT CGC AGG CTG GCA GGC ATT CTG G and its reverse complement.

FRB-BirA-pcDNA4-Tet and FKBP12-AP-pcDNA4-Tet mammalian expression plasmids

Both constructs were created by direct digestion from the corresponding pcDNA3 plasmids followed by ligation into the HindIII and EcoRI sites of pcDNA4-Tet.

NLS-FKBP-AP(-3)-pcDNA4-Tet mammalian expression plasmid

The Myc-FKBP-AP(-3) gene was PCR-amplified from the corresponding pcDNA3 plasmid with the primers: 5′- GGT ATG GCT AGC ATG ACT GGT G and 5′-TTT GAA TTC TTA CAC CTT GCG CTT CTT CTT GGG CAC CTT GCG CTT CTT CTT GGG CAC CTT GCG CTT CTT CTT GGG GCC GCC CTC GAT CTT CTG GGC CTC G to introduce an NheI site at the N-terminus, and three tandemly repeated nuclear localization signals (NLS, protein sequence PKKKRKVPKKKRKVPKKKRKV) and an EcoRI site at the C-terminus. The PCR product was ligated into the NheI and EcoRI sites of pcDNA4-Tet (Invitrogen).

NLS-FRB-BirA-pcDNA4-Tet mammalian expression plasmid

The HA-FRB-BirA gene was PCR-amplified from the corresponding pCDNA3 plasmid with the primers: 5′- GGT ATG GCT AGC ATG ACT GGT G and 5′- TTT GAA TTC TTA CAC CTT GCG CTT CTT CTT GGG CAC CTT GCG CTT CTT CTT GGG CAC CTT GCG CTT CTT CTT GGG GCC TTT TTC TGC ACT ACG CAG GG to introduce an NheI site at the N-terminus, and three tandemly repeated nuclear localization signals (NLS, protein sequence PKKKRKVPKKKRKVPKKKRKV) and an EcoRI site at the C-terminus. The PCR product was ligated into the NheI and EcoRI sites of pcDNA4-Tet (Invitrogen).

Expression and purification of recombinant FKBP12, FRB, and HP1 fusion proteins

E. coli BL21(DE3) cells transformed with one of the pET21 expression plasmids were amplified in LB media supplemented with 100 μg/mL ampicillin at 37 °C until OD_{600} 0.9. Protein expression was induced with 100 μg/mL IPTG for 3 hours at 30 °C. Cells were harvested by centrifugation and the pellet was resuspended in lysis buffer (50 mM Tris base, 300 mM NaCl, pH 7.8) containing 2.5 mM PMSF and protease inhibitor cocktail (Calbiochem). Cells were lysed by ultrasonic treatment (six 15-second bursts, with 1 minute of cooling to 4 °C between bursts). The extract was cleared by centrifugation (17,700 g, 10 minutes, 4 °C) and the His_6-tagged protein was purified using Ni-NTA agarose (Qiagen). Fractions were analyzed by 12–16% SDS-PAGE followed by Coomassie staining. Fractions containing protein were pooled and dialyzed against PBS pH 7.4. Protein concentrations were measured using the BCA assay (Pierce) with BSA as the reference standard. Purified proteins were stored in PBS pH 7.4 at −80 °C for up to 6 months.

***In vitro* FKBP12-FRB interaction assay**

Purified FRB-BirA and FKBP12-AP(-3) (or FKBP12-AP) proteins were combined at either 1 μM or 100 nM in PBS supplemented with 50 μM biotin, 5 mM $MgCl_2$, and 1 μM rapamycin. Parallel reactions were run with rapamycin omitted. Reactions were initiated with addition of 5 mM ATP, and incubated at room temperature for 1 minute (for 1 μM protein reactions) or 5 minutes (for 100 nM protein reactions),

before quenching with 0.1 volume of 500 mM EDTA pH 8.0. Reaction products were analyzed by streptavidin blotting, as described below.

Western blotting

Cell lysates or *in vitro* reaction products were combined with SDS-PAGE loading buffer containing 2-mercaptoethanol, boiled, run on a 16% SDS-PAGE gel, and transferred to nitrocellulose membrane. For lysate analysis, approximately 10^5 cells were loaded in each lane. For analysis of *in vitro* reactions, 20 µL of each reaction was loaded per lane. Membranes were blocked with TBS + 0.05% Tween 20 (TBS-T) and 3% BSA for 1 hour. For streptavidin blotting, the membrane was incubated with 0.67 µg/mL ImmunoPure streptavidin-horseradish peroxidase conjugate (Pierce) in TBS-T + 3% BSA for 40 minutes, followed by four 5-minute washes in TBS-T. For anti-myc blotting, the membrane was first incubated with 10 µg/mL mouse anti-c-myc antibody (Calbiochem) in TBS-T + 3% BSA for 45 minutes, then washed three times. The membrane was then incubated with anti-mouse horseradish peroxidase conjugate (Bio-Rad) in TBS-T + 3% BSA at 1:2000 dilution for 45 minutes, followed by three rounds of 5 minute washes with TBS-T. Blots were developed with Supersignal West Pico or Femto substrate (Pierce), and images were taken on a ChemiImager 5500 instrument (Alpha Innotech). The intensity of each band was quantified using AlphaEase FC version 3.2.2 software (Alpha Innotech).

Comparison of FKBP12-AP′ truncation mutants in live cells

HEK 293T cells were grown in DMEM with 10% FBS, 50 units/mL penicillin, and 50 µg/mL streptomycin, and co-transfected with FRB-BirA-pcDNA3 and FKBP12-AP′-pcDNA3 in a 1:1 ratio using Lipofectamine 2000 (Invitrogen). After transfection, the cells were grown in culture media depleted of biotin. Biotin-free media was prepared by incubating with streptavidin agarose slurry (Novagen) (100 µL agarose per 1 mL media) for 30 minutes at room temperature, followed by filtering. After 48 hours of transfection, the appropriate amount of rapamycin (prepared as a 1 mM stock in DMSO) was added to the culture media to a final concentration of 100 nM, and the cells were incubated for an additional 1 hour at 37 °C. Thereafter, the media was replaced with pre-warmed PBS-Mg

and 10 μM biotin for 1 minute. The cells were then washed three times with ice-cold PBS-Mg to stop the biotinylation reaction. To analyze biotinylated proteins, cells were resuspended and lysed in triton lysis buffer (1% Triton X-100, 150 mM NaCl, 5 mM EDTA, 20 mM Tris pH 7.4, 2 mM PMSF, and 1% protease inhibitor cocktail (Calbiochem)) for 20 minutes on ice. The supernatant was either stored at −80 °C or analyzed immediately by western blotting as described above.

Analysis of detection specificity in live cells

HEK 293T cells were grown in DMEM with 10% FBS, 50 units/mL penicillin, and 50 μg/mL streptomycin, and co-transfected with FRB-BirA-pcDNA3 and FKBP12-AP(-3)-pcDNA3 in a 1:1 ratio using Lipofectamine 2000 (Invitrogen). For the negative controls, the appropriate plasmids were co-transfected to express the FRB-BirA(K183R) or FKBP12-AP(-3)-Ala mutants. After transfection, the cells were grown in culture media supplemented with 50 μg/mL of soluble streptavidin to deplete the biotin in media. After 12–16 hours of transfection, the cells were incubated for 3 hours at 37 °C in the presence or absence of 100 nM rapamycin. The media was then replaced with pre-warmed PBS-Mg and 50 μM biotin for 1 minute. Thereafter, the cells were washed, lysed, and analyzed as above.

Measurement of k_{cat} and K_M for AP(-3) biotinylation by BirA

Kinetic measurements were performed by incubating different concentrations of the HP1-AP(-3) fusion protein (50–1000 μM) with 3.3 μM BirA, 1 mM biotin, 400 nM ^{3}H-biotin, 5 mM magnesium acetate, and 1% BSA in 50 mM bicine buffer pH 7.4. The reactions were initiated with addition of 4 mM ATP to the pre-warmed (30 °C) mixture. The reactions were incubated at 30 °C, and 10 μL aliquots were removed every 30 minutes and quenched with 10% TCA (final concentration). The quenched reaction mixtures were then diluted to give a final concentration of 50 μM HP1-AP(-3) and spotted onto phosphocellulose circles that had been pretreated with 1 mM biotin and 10% TCA. The circles were then washed twice with cold 10% TCA, once with ethanol, and dried. All the phosphocellulose washing steps were performed using a vacuum manifold

(Millipore). The acid-insoluble radioactivity spotted on the circles was measured by liquid scintillation counting. Negative control reactions omitting ATP were used for background correction. Kinetic data points were plotted and fitted against the Michaelis-Menten equation using ORIGIN software.

Rapamycin dose-response curve

This test was performed as described above for the live-cell comparison of the FKBP12-AP′ constructs except that all cells were co-transfected with FRB-BirA-pcDNA3 and FKBP12-AP(-3)-pcDNA3; and the cells were treated with varying amounts of rapamycin ranging from 0–330 nM.

Titration of protein concentration for the detection of cellular PPIs

T-REx™-293 cells (Invitrogen) were grown in DMEM with 10% FBS, 50 units/mL penicillin, 50 μg/mL streptomycin, and 5 μg/mL blasticidin and co-transfected with FRB-BirA-pcDNA4-Tet and FKBP12-AP(-3)-pcDNA4-Tet in a 1:1 ratio using Lipofectamine 2000 (Invitrogen). After transfection, the cells were grown in culture media supplemented with 50 μg/mL of soluble streptavidin to deplete the biotin in media and with varying concentrations of tetracycline (0–1 μg/mL). After 16–24 hours of transfection, the cells were incubated for 3 hours at 37 °C in the presence or absence of 100 nM rapamycin. The media was then replaced with pre-warmed PBS-Mg and 50 μM biotin for 1 minute. Thereafter, the cells were washed, lysed, and analyzed as for the Comparison of FKBP12-AP′ truncation mutants in live cells.

Quantification of the extent of biotinylation upon PPI detection in cells

T-REx™-293 cells (Invitrogen) were transfected, treated, and lysed as above. Cell extracts were incubated overnight at 4 °C in the presence or absence of 2.8 μM soluble streptavidin, and the binding reactions were run in a 12% SDS-PAGE without boiling the samples (to preserve the biotin-streptavidin bond). Analysis was done by anti-myc blotting as described above.

Imaging protein-protein interactions in the cytosol of mammalian cells

HEK 293T were grown and transfected as described for the specificity test above. After the biotin treatment, the cells were fixed with 3.7% paraformaldehyde for 15 minutes at room temperature followed by permeabilization with methanol for 6 minutes at –20 °C. The samples were washed with PBS-Mg and blocked with PBS-Mg + 3% BSA for 1 hour at room temperature with rocking. Biotin detection was performed by incubating the cells with 5 μg/mL streptavidin-Alexa Fluor 568 conjugate (Invitrogen) in PBS-Mg + 3% BSA + 0.03% Tween20 for 1 hour at 4 °C followed by three 10-minute washes with PBS-Mg + 0.1% Tween20. Detection of the HA epitope in the FRB-BirA protein was performed by incubating the cells with 5 μg/mL rabbit anti-HA antibody (Rockland Immunochemicals) followed by 6.7 μg/mL anti-rabbit antibody conjugated to Alexa Fluor 488 (Invitrogen). After washing, the cells were imaged in PBS-Mg with a Zeiss Axiovert 200M inverted epifluorescence microscope using a 63x oil-immersion lens and a Cascade II:512 camera (Photometrics). Alexa Fluor 568 (560/20 excitation, 585 dichroic, 605/30 emission), Alexa Fluor 488 (495/20 excitation, 515 dichroic, 530/30 emission), and differential interference contrast (DIC) images (630/10 emission) were collected and analyzed using Slidebook software (Intelligent Imaging Innovations). Fluorescence images were normalized to the same intensity range. Acquisition times ranged from 10–100 milliseconds.

Imaging protein-protein interactions in the nucleus of mammalian cells

The experiment was performed as described for the cytosolic proteins except that T-REx™-293 cells were used. T-REx™-293 cells (Invitrogen) were grown in DMEM with 10% FBS, 50 units/mL penicillin, 50 μg/mL streptomycin, and 5 μg/mL blasticidin and co-transfected with FRB-BirA-pcDNA4-Tet and FKBP12-AP(-3)-pcDNA4-Tet in a 1:1 ratio using Lipofectamine 2000 (Invitrogen). After transfection, protein expression was induced by addition of 0.4 μg/mL tetracycline to the cell media. Cells were then treated with rapamycin and biotin, fixed, stained, and imaged as above. Additional detection of the myc epitope in the FKBP-AP(-3) protein was performed by incubating

the cells with 5 μg/mL mouse anti-c-myc antibody (Calbiochem) followed by 6.7 μg/mL anti-rabbit antibody conjugated to Alexa Fluor 568 (Invitrogen). Biotin detection was performed as above except that streptavidin-Alexa Fluor 647 conjugate was used instead of streptavidin-Alexa Fluor 568 conjugate. The following filters were used for Alexa Fluor 647 detection: 630/10 excitation, 660 dichroic, 680/30 emission.

References

1. Remy,I. & Michnick,S.W. Mapping biochemical networks with protein-fragment complementation assays. *Methods Mol. Biol.* **261**, 411-426 (2004).

2. Beckett,D., Kovaleva,E. & Schatz,P.J. A minimal peptide substrate in biotin holoenzyme synthetase-catalyzed biotinylation. *Protein Sci.* **8**, 921-929 (1999).

3. Chen,I., Howarth,M., Lin,W. & Ting,A.Y. Site-specific labeling of cell surface proteins with biophysical probes using biotin ligase. *Nat. Methods* **2**, 99-104 (2005).

4. Howarth,M., Takao,K., Hayashi,Y. & Ting,A.Y. Targeting quantum dots to surface proteins in living cells with biotin ligase. *Proc. Natl. Acad. Sci. U. S. A* **102**, 7583-7588 (2005).

5. de Boer,E. *et al.* Efficient biotinylation and single-step purification of tagged transcription factors in mammalian cells and transgenic mice. *Proc. Natl. Acad. Sci. U. S. A* **100**, 7480-7485 (2003).

6. Diamandis,E.P. & Christopoulos,T.K. The biotin-(strept)avidin system: principles and applications in biotechnology. *Clin. Chem.* **37**, 625-636 (1991).

7. Howarth,M. *et al.* A monovalent streptavidin with a single femtomolar biotin binding site. *Nat. Methods* **3**, 267-273 (2006).

8. Howarth,M. *et al.* Monovalent, reduced-size quantum dots for imaging receptors on living cells. *Nat. Methods* (2008).

9. Howarth,M. & Ting,A.Y. Imaging proteins in live mammalian cells with biotin ligase and monovalent streptavidin. *Nat. Protoc.* **3**, 534-545 (2008).

10. McMahon,R.J. Biotin in metabolism and molecular biology. *Annu. Rev. Nutr.* **22**, 221-239 (2002).

11. Chen,J., Zheng,X.F., Brown,E.J. & Schreiber,S.L. Identification of an 11-kDa FKBP12-rapamycin-binding domain within the 289-kDa FKBP12-rapamycin-

associated protein and characterization of a critical serine residue. *Proc. Natl. Acad. Sci. U. S. A* **92**, 4947-4951 (1995).

12. Bierer,B.E. *et al.* Two distinct signal transmission pathways in T lymphocytes are inhibited by complexes formed between an immunophilin and either FK506 or rapamycin. *Proc. Natl. Acad. Sci. U. S. A* **87**, 9231-9235 (1990).

13. Banaszynski,L.A., Liu,C.W. & Wandless,T.J. Characterization of the FKBP.rapamycin.FRB ternary complex. *J. Am. Chem. Soc.* **127**, 4715-4721 (2005).

14. Choi,J., Chen,J., Schreiber,S.L. & Clardy,J. Structure of the FKBP12-rapamycin complex interacting with the binding domain of human FRAP. *Science* **273**, 239-242 (1996).

15. Galarneau,A., Primeau,M., Trudeau,L.E. & Michnick,S.W. Beta-lactamase protein fragment complementation assays as in vivo and in vitro sensors of protein protein interactions. *Nat. Biotechnol.* **20**, 619-622 (2002).

16. Luker,K.E. *et al.* Kinetics of regulated protein-protein interactions revealed with firefly luciferase complementation imaging in cells and living animals. *Proc. Natl. Acad. Sci. U. S. A* **101**, 12288-12293 (2004).

17. Pelletier,J.N., Campbell-Valois,F.X. & Michnick,S.W. Oligomerization domain-directed reassembly of active dihydrofolate reductase from rationally designed fragments. *Proc. Natl. Acad. Sci. U. S. A* **95**, 12141-12146 (1998).

18. Remy,I. & Michnick,S.W. Clonal selection and in vivo quantitation of protein interactions with protein-fragment complementation assays. *Proc. Natl. Acad. Sci. U. S. A* **96**, 5394-5399 (1999).

19. Rossi,F., Charlton,C.A. & Blau,H.M. Monitoring protein-protein interactions in intact eukaryotic cells by beta-galactosidase complementation. *Proc. Natl. Acad. Sci. U. S. A* **94**, 8405-8410 (1997).

20. Wehrman,T., Kleaveland,B., Her,J.H., Balint,R.F. & Blau,H.M. Protein-protein interactions monitored in mammalian cells via complementation of beta -lactamase enzyme fragments. *Proc. Natl. Acad. Sci. U. S. A* **99**, 3469-3474 (2002).

21. Chen,I., Choi,Y.A. & Ting,A.Y. Phage display evolution of a peptide substrate for yeast biotin ligase and application to two-color quantum dot labeling of cell surface proteins. *J. Am. Chem. Soc.* **129**, 6619-6625 (2007).

22. Yao,F. *et al.* Tetracycline repressor, tetR, rather than the tetR-mammalian cell transcription factor fusion derivatives, regulates inducible gene expression in mammalian cells. *Hum. Gene Ther.* **9**, 1939-1950 (1998).

23. Chapman-Smith,A. & Cronan,J.E., Jr. The enzymatic biotinylation of proteins: a post-translational modification of exceptional specificity. *Trends Biochem. Sci.* **24**, 359-363 (1999).

24. Weaver,L.H., Kwon,K., Beckett,D. & Matthews,B.W. Corepressor-induced organization and assembly of the biotin repressor: a model for allosteric activation of a transcriptional regulator. *Proc. Natl. Acad. Sci. U. S. A* **98**, 6045-6050 (2001).

25. Remy,I., Ghaddar,G. & Michnick,S.W. Using the beta-lactamase protein-fragment complementation assay to probe dynamic protein-protein interactions. *Nat. Protoc.* **2**, 2302-2306 (2007).

26. Demidov,V.V. *et al.* Fast complementation of split fluorescent protein triggered by DNA hybridization. *Proc. Natl. Acad. Sci. U. S. A* **103**, 2052-2056 (2006).

27. Stefan,E. *et al.* Quantification of dynamic protein complexes using Renilla luciferase fragment complementation applied to protein kinase A activities in vivo. *Proc. Natl. Acad. Sci. U. S. A* **104**, 16916-16921 (2007).

Part II : New methodology for site-specific protein labeling in live cells

Chapter 5 : Introduction. Methodologies for site-specific protein labeling in live cells

Introduction

Part I of this thesis dealt with the development of new reporters of protein-protein interactions in cells. We described how protein interaction networks lie at the core of most cellular processes, and how these interactions are tightly regulated both in space and in time. For this reason, elucidating the distribution, dynamics, and chemical environment of individual proteins is critical, and often comes before any analysis of its interacting partners. Part II of this thesis, thus focuses on the development of new reporters to monitor protein localization and trafficking.

Optical and, in particular, fluorescence imaging has revolutionized cell biology by enabling the visualization and monitoring of proteins inside living cells with high temporal and spatial resolution. These studies often rely on recombinant fusions of the target protein to a variant of the green fluorescence protein (GFP). Although fluorescent protein fusions have shed light into the localization and function of many cellular proteins[1], their use has several limitations. Probably the major concern with FPs is their large size, which has the potential to interfere with the function of the protein under study. Indeed, GFP fusions have been shown to promote aggregation of several membrane-targeted proteins[2], to cause retention in the endoplasmic reticulum[3], and to alter the dynamics of receptor recycling[4,5]. Fluorescent proteins are also not bright or photostable enough to perform single-molecule imaging in live cells, and they cannot easily distinguish different protein subpopulations, such as cell-surface and intracellular (unless photoactivatable FPs are used). Additionally, fluorescent proteins are only passive reporters, which can report on protein localization but do not provide information on its biochemical activity or on its age. Finally, FPs are in general only amenable to optical imaging but not to other forms of imaging such as electron microscopy, which can provide higher spatial resolution, or magnetic resonance imaging and positron emission tomography, which can extent imaging studies to living animals.

For this reason, our lab and others are trying to develop new protein labeling methods that circumvent the limitations posed by GFP. And we are doing so by combining the enhanced imaging properties and diverse functionalities of chemical (non-genetically encoded) probes, with the power of genetic targeting. Chemical probes, such as small-molecules organic fluorophores, inorganic quantum dots, or photoaffinity probes

overcome many of the limitations of GFP. For example, organic fluorophores are smaller and brighter than FPs, whereas photoaffinity probes can be used to study protein-protein interactions. Quantum dots, on the other hand, although bigger than organic probes, are far brighter than FPs and do not photobleach; and are therefore amenable for single-molecule imaging in live cells. Although these probes are extensively used in *in vitro* studies, their application to cellular imaging has been hampered by the shortage of robust methods to target them with high specificity to the protein of interest. Because chemical probes are non-genetically encoded, they cannot be simply genetically fused to the target protein, as it is done with FPs.

Therefore, in recent years, many new methods have been developed that try to bridge the exquisite specificity of genetic targeting with the powerful imaging properties of chemical probes. In the next sections, we describe some of these methods.

Traditional ways of targeting chemical probes to live cells: antibodies and receptor ligands

As described in Chapter 1, immunostaining is one of the most common ways of imaging the localization of endogenous cellular proteins. In this method, antibodies against the target proteins are conjugated to a chemical probe (typically a small organic fluorophore) and used to stain a biological sample. Although immunostaining has been traditionally employed to stain fixed samples, it can also be used in living cells to label cell-surface proteins. Indeed, immunofluorescence (IF) has been extensively used to study the localization and trafficking of cellular receptors. Receptors are cell-surface proteins that mediate communication between the cytoplasm and the external environment through the recognition of signaling molecules and uptake of nutrients. Because many aspects of receptor function are controlled by spatial localization and temporal dynamics, imaging is a powerful method to study the behavior of membrane receptors. Antibody targeting of chemical probes, however, still shares one of the main disadvantages of GFP; that of its large size. Because antibodies are large (~ 150 kD), they often interfere with receptor functions, such as ligand binding and protein

oligomerization. Additionally, because of the bivalent nature of their binding, antibodies can crosslink receptors, changing their trafficking and causing unintentional activation[6,7]. Finally, because binding is non-covalent, antibodies often dissociate from their target after receptor internalization, thus leading to artifacts in studies of receptor recycling[8]. Although the use of monovalent, smaller Fab fragments minimizes the problems of steric interference and protein crosslinking, they typically display off rates of 10^{-4}–10^{-3} s^{-1} (i.e., half-lives of minutes to few hours), which results in poor binding and rapid signal degradation[9,10].

Another common method for imaging receptors is to use fluorescently-tagged ligands, such as epidermal growth factor (EGF)[11], transferrin[12], and low-density lipoprotein (LDL)[13]. However, because ligands also dissociate from their receptors after internalization, this method does not allow tracking of receptor fate and recycling. In addition, the trafficking behavior of unactivated, ligand-free receptor cannot be studied.

Genetic targeting of chemical probes in live cells

Many new protein labeling methods have been developed in recent years[14,15], and they all share a common underlying goal: to combine the excellent specificity of genetic targeting with the diverse photophysical properties of chemical probes. The basic principle is hence to genetically fuse the target protein to a recognition element (i.e., a peptide or a full-length protein), which is capable of recruiting a chemical probe. Figure 5-1 depicts the three most common approaches that have been devised.

In the first approach (Figure 5-1A), the protein of interest is fused to a peptide sequence, which then recruits the small-molecule. In this approach, because the binding of the probe to the protein is not covalent, the usefulness of the method depends on a tight binding affinity between the peptide and the probe. Examples of this are the His_6-NTA[16-18], the polyaspartate[19-21], and the tetracysteine (FlAsH)[22,23] methodologies. In the second approach (Figure 5-1B), the target protein is fused to a full-length protein, instead of a peptide, which can bind to the small-molecule covalently or non-covalently. Examples of non-covalent attachment using this approach are the methodologies based

on the FKBP12[24] and DHFR[25,26] proteins, whereas covalent ligation is employed in the SNAP-tag[27-29] and HaloTag[30] techniques.

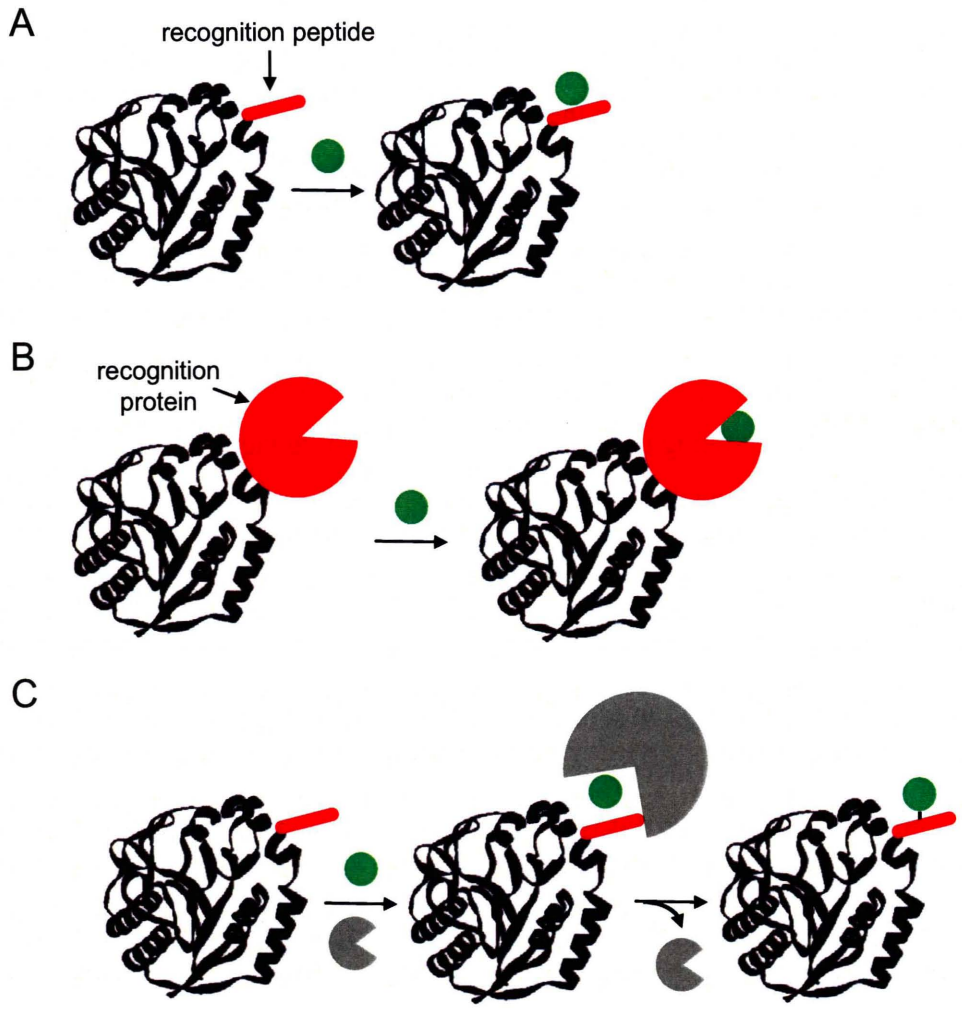


Figure 5-1: Strategies for chemical targeting of small molecules to proteins in live cells. (**A**) Small-molecule labeling using peptide tags. The protein of interest (black, ribbon format) is genetically fused to a peptide tag (e.g., tetracysteine or hexahistidine; red) that is able to non-covalently bind to a small-molecule (e.g., biarsenical- or NTA-fluorophore conjugates; green dot) (**B**) Small-molecule labeling using protein tags. The protein of interest is genetically fused to a protein tag (e.g., DHFR or hAGT; red) that is able to bind non-covalently or covalently to a small molecule (e.g., methotrexate or O^6-benzylguanine conjugate; green dot). (**C**) Small molecule labeling using enzyme-mediated covalent attachment to a peptide tag. An enzyme (e.g., phosphopantetheine transferase or biotin ligase; grey) labels a genetically-fused peptide tag (e.g., S6 or AP; red) with the small molecule (green dot) directly or with a functional handle to which the small molecule probe ligates through a chemoselective reaction.

A survey of the techniques employing the first two approaches reveals that a general trade-off exists between labeling specificity and tag size. Protein-based tags, such

as SNAP- and Halo-tag generally give higher labeling specificity than peptide tags, such as FlAsH. However, protein tags have increased potential to interfere with protein folding, trafficking, and activity, as we have described for GFP. Our lab and others have tried to bridge the requirements of small tag size and high labeling specificity, by making use of enzyme ligases. Whereas in the two previous approaches the labeling specificity was solely determined by the interactions established during the bimolecular reaction between the recognition element and the probe (Figure 5-1A & B), in the third approach, an enzyme is employed to mediate the conjugation of the probe to the target sequence (Figure 5-1C). By capitalizing on the intrinsic sequence specificity of enzymes such as biotin ligase or phosphopantetheine transferase, highly specific probe conjugation can be achieved, without sacrificing the small size of the directing tag. Examples of this third class of techniques are the methods based on the phosphopantetheine transferases (Sfp and AcpS)[31-33], transglutaminase[34], and biotin ligase[35-37] enzymes.

In the next sections, we will describe each one of these technique in more detail. We will specifically emphasize their weaknesses and strengths with regard to the labeling specificity, tag size, stability of probe-protein linkage, and their potential to be applied in any subcellular compartment (summarized in Table 5-1, at the end of the chapter). As it will be inferred from the analysis, there currently exists no methodology capable of attaching small-molecule probes to intracellular proteins modified by just a small peptide tag, with high specificity and without toxic side-effects. It was this lack of an ideal method, which motivated our development of a new site-specific protein labeling methodology based on the lipoic acid ligase (Chapters 6, 7, and 8)[38].

Small-molecule labeling using peptide recognition sequences

In this approach, the protein of interest is genetically fused to a peptide tag that is capable of binding to the desired chemical probe (Figure 5-1A). Whereas the use of a peptide is generally preferred because of its small size, attention must be paid to the affinity of the binding interaction such that a good signal-to-noise ratio can be achieved of specific peptide-probe binding over background nonspecific interactions of the probe with other cellular components.

Probably the most successful demonstration of this type of methodologies is the FlAsH technique developed in the laboratory of Roger Tsien (U.C. San Diego)[39]. In this method, the target protein is fused to a 6–16 amino acid peptide that contains four cysteines (e.g., CCPGCC) and binds very tightly to a biarsenical dye (the FlAsH and ReAsH probes), which becomes intensely fluorescent only upon binding to the tetracysteines. One of the main advantages of this methodology is that the probes are membrane-permeant, and thus can be used to label intracellular proteins. Most importantly, what distinguishes this method from all the others in this section, is its high affinity binding interaction ($K_d \sim 2$–4 pM), such that the signal does not deteriorate over time. Recent improvements to the original methodology have also greatly enhanced its utility. For example, new biarsenical compounds that span the visible spectrum[40], as well as probes useful for electron microscopy imaging[22], and singlet oxygen generators that inactivate protein function[41] have been developed. Unlike most of the other methods described in this chapter, the combination of tight binding with membrane permeant probes has allowed the FlAsH methodology to be widely applied in cell biology, where it has already begun to make significant contributions[22,42-45]. However, the FlAsH methodology still has some associated problems that limit its utility. The weak affinity of the biarsenical dyes for mono- and di-cysteines, which are present in most intracellular proteins, results in high background labeling unless an antidote such as 1, 2-ethanedithiol is used. Unfortunately, the use of antidotes translates into a long labeling procedure. For this reason, the method usually works best on overexpressed proteins[46]; although new tetracysteine sequences that bind with higher affinity to the biarsenical probes may circumvent this limitation[40,47]. Other limitations remain, however, such as the arsenic toxicity and the requirement for the cysteines to be in reduced form for probe attachment, which prevents the FlAsH technique from labeling proteins in the secretory pathway and at the cell surface.

Other peptide tags that bind fluorophores directly include the hexahistidine, polyaspartate and Texas Red binding peptides. The 38-amino acid Texas Red binding peptide was evolved by Nolan and co-workers to bind with high-affinity ($K_d \sim 25$ pM) to the Texas Red fluorophore, and it was shown to label GFP expressed at the surface of mammalian cells[48]. Even though the measured *in vitro* binding affinity was relatively

high, the use of Texas Red presents a problem in live cells because of its intrinsic affinity for the mitochondrial membrane, which results in high background.

Another labeling method is based on the binding of oligohistidine sequences to nickel (II) nitrilotriacetic acid (Ni-NTA) complexes[16,18,49]. However, due to the low affinity of the complex (initial $K_d \sim 1–20$ μM), this method suffers badly from probe detachment (i.e., signal deterioration over time) and Ni^{2+} toxicity. These limitations were partially addressed by the development of a fluorescein dizinc compound that binds with nanomolar affinity to hexahistidine sequences[50], but the method is still restricted to cell-surface proteins. An analogous approach has also been developed to label cell-surface proteins, which involves binding of an oligoaspartate (8–16 amino acids) to a multinuclear Zn(II) complex (i.e., two Zn(II) atoms complexed to four 2,2′-dipicolylamine within a tyrosine scaffold)[19,21]. As in the case of the oligohistidine technique, the original method suffered from severe probe dissociation. The problem was later solved by Nonaka and co-workers, who added a chloroacetamide functionality to the Zn(II) complex[20]. Upon binding to the oligoaspartate-tagged protein, the chloroacetamide reacts with a protein surface cysteine thus forming a covalent bond that prevents probe dissociation[20]. Imperiali and co-workers have also developed genetically encodable peptide tags that bind with nanomolar affinity to lanthanide ions[51]. The most interesting characteristic of this approach is that lanthanide ions, such as Tb^{3+}, are sensitized by the tryptophan residue present in the lanthanide-binding peptide, which results in long-life luminescence emissions. This type of emission (in contrast to fluorescence) is particularly well suited for x-ray and NMR structure determinations[52,53] and for studying protein-protein interactions using luminescence resonance energy transfer[54]. Lanthanide-labeling has not been used yet in live cells, probably because of the need for phototoxic ultraviolet excitation of the lanthanide ion. Recently, this problem has been addressed by Reynolds and colleagues, who reported that replacement of tryptophan by unnatural amino acids (i.e., carbostyril and acridone) allowed for longer wavelength excitation[55]. However, the maximum excitation achieved was 390 nm, where phototoxicity is still an issue.

Finally, streptavidin-binding peptides and bungarotoxin-binding peptides have been used as tags for site-specific protein labeling[56]. However, the large size of the streptavidin- and bungarotoxin-probe conjugates partially defeats the purpose of using a

small peptide tag. Additionally, this method is restricted to labeling membrane proteins because both streptavidin and bungarotoxin are membrane-impermeant.

As it can be readily inferred from the above description, the main common limitation of all these methodologies is the high background labeling caused by the relatively low binding affinity of the probe towards the peptide, when compared to other non-specific interactions (e.g., monothiols in the case of the biasernical dyes or the mitochondrial membrane in the case of Texas Red). In order to overcome this limitation, other methods have been developed, which use a full-length protein instead of a peptide recognition element.

Small-molecule labeling using protein recognition sequences

The two main examples of non-covalent binding of chemical probes to protein recognition sequences rely on the high affinity interaction of a small-molecule substrate to either *E. coli* dihydrofolate reductase (DHFR) or the F36V mutant of the FK506 binding protein 12 (FKBP12 F36V). DHFR binds to methotrexate and trimethoprim with picomolar dissociation constants[57], and this interaction affinity is retained when the inhibitors are conjugated to different fluorophores. Importantly, because both, methotrexate and trimethoprim conjugates are membrane-permeant, this is one of the few site-specific labeling methodologies that can target intracellular proteins. In their first report, Cornish and co-workers labeled nuclear and plasma-membrane localized DHFR fusion proteins in DHFR-deficient Chinese hamster ovary (CHO) cells using a Texas Red methotrexate conjugate[25]. The need to work in DHFR-deficient cell lines was later avoided by using trimethoprim conjugates, which display a larger discrimination for binding to *E. coli* DHFR instead of its mammalian counterpart[26,58]. In an analogous approach, the binding interaction of FKBP12 F36V to a fluorescein-conjugate of the synthetic ligand for FKBP12 (SLF′) was exploited by the Nolan laboratory to label membrane-targeted proteins fused to FKBP12 F36V in living mammalian cells[24]. FKBP12 has the advantage over DHFR, that it has a > 1000-fold preference for the F36V mutant (not endogenous to mammalian cells) over wild-type FKBP12. Labeling with the DHFR and FKBP12 tags is fast and specific, but, like the FPs and antibodies, these

proteins are large (157 and 108 amino acids, respectively). Furthermore, even with such high interaction affinities, the methods still suffer from signal deterioration over time. For example, the Texas Red methotrexate conjugate dissociates from the DHFR fusion proteins after just one hour[25].

In order to eliminate the problem of probe dissociation, other methods have been developed that exploit the covalent attachment of a chemical probe to a protein or enzyme tag. The first example of this strategy was reported by the Johnsson laboratory and it is based on the irreversible and specific reaction of the human enzyme O^6-alkylguanine transferase (hAGT) with O^6-benzylguanine derivatives, leading to the transfer of the synthetic probe to the reactive Cys 145 in hAGT[28]. The authors termed the fused hAGT as the SNAP-tag, owing to the high speed and covalent nature of the labeling. Using the SNAP-tag, Kepler and co-workers achieved rapid and specific labeling of nuclear, cytosolic and membrane proteins fused to hAGT in *E. coli*, yeast, and mammalian cells[28,29]. Although, as in the case of DHFR, the method was at first restricted to hAGT-deficient cells lines, evolution of an orthogonal mutant later circumvented the problem and eliminated background staining[59]. Recently, the development of an orthogonal pair based on the reaction of a mutant hAGT with O^6-propargylguanine derivatives has enabled two-color labeling of intracellular protein in living mammalian cells[27]. This orthogonal pair has been termed the CLIP-tag. Following a similar approach, Promega has developed a protein labeling method based on a bacterial haloalkane dehalogenase, engineered to form covalent enzyme-substrate adducts. The method, called HaloTag™, has been used to covalently attach various chloroalkane-bearing fluorophores[30] and quantum dots[60] to cell-surface proteins fused to the dehalogenase enzyme. Finally, the fungal enzyme cutinase has also been exploited to covalently attach organic fluorophores and quantum dots to the integrin lymphocyte function-associated antigen-1 (LFA-1) on the surface of living cells[61]. All three methods (SNAP-tag, HaloTag, and cutinase) are characterized by high specificity, fast labeling reactions (10–30 minutes), and covalent probe attachment, and thus represent an improvement over the DHFR and FKBP12 F36V labeling methods described above. Unfortunately, they all still share the common drawback of employing large protein tags (182, 296, and 213

amino acid residues, respectively), which can potentially interfere with the function of the protein under study.

Enzyme-mediated ligation of small-molecules to peptide recognition sequences

The two previous approaches rely on the most straightforward form of joining a chemical probe with a peptide or protein recognition sequence: a binding interaction. However, it has become clear that, when a mere bimolecular interaction is established, there is necessarily a trade-off between the size of the interactors (e.g., peptide *versus* protein) and the specificity of the interaction. If a large protein tag is used, the amount of interactions created between the probe and the tag are larger, thus resulting in more specific labeling. To overcome this limitation, our lab and others have developed an alternative approach, which attempts to increase specificity while using a small peptide tag. The approach involves using an enzyme to mediate the labeling (Figure 5-1C). By using an enzyme, the interaction surface area between the probe and the peptide increases, thus resulting in more specific labeling. Moreover, enzyme-mediated labeling results in the covalent attachment of the probe to the peptide recognition, which enables long-term imaging of the same molecule. Enzyme-mediated labeling should also bring about faster and more sensitive labeling owing to the enforced proximity between the probe and the peptide within the enzyme active site.

Phosphopantetheine transferases (PPTases)were the first enzymes exploited for enzyme-mediated protein labeling[31,32,62]. PPTases catalyze the transfer of 4′-phosphopantethine from coenzyme A (CoA) to a serine residue of an acyl or peptidyl carrier protein (ACP and PCP). One of the main advantages of the use of PPTases for protein labeling is the fact that they are tolerant of a wide range of substitutions at the terminal thiol of the cofactor CoA, allowing the transfer of diverse fluorophores or other chemical probes to ACP- or PCP-tagged proteins. Additionally, although the original methodology relied on protein tags (ACP and PCP are both ~ 80 amino acid long), the Yin lab recently evolved 12-mer peptide tags that are modified by PPTases with similar kinetics to the original full-length proteins, greatly enhancing the utility of the methodology[33]. However, while the PPTase methodology results in fast, specific, and

covalent labeling, the highly-charged CoA derivatives are membrane impermeant, which limits its utility to cell-surface proteins. Nonetheless, the PPTase methodology is, together with the FlAsH technique, one of the most widely used methods to genetically target chemical probes to protein in living cells. The method has been used for fluorophore tagging of the agglutinin receptor Aga2p in yeast and the human G-protein coupled receptor NK1in mammalian cells[63], for FRET studies of the transferrin receptor[64], and for single-molecule studies of the odorant receptor trafficking[65].

Our lab has exploited the use of two different enzymes for protein labeling: biotin ligase and transglutaminase. *E. coli* biotin ligase (BirA) catalyzes the ATP-dependent ligation of biotin to a lysine side-chain within a 15-amino acid acceptor peptide substrate (AP). We have used BirA to label AP-tagged cell-surface proteins with fluorophores and quantum dots conjugated to streptavidin[36,37,66,67]. Recently, the methodology has also been extended to multicolor imaging[68]. The labeling is extremely specific and fast (i.e., 5-10 minutes), but, as previously noticed, the use of the large streptavidin protein (~ 56 kD) to detect the ligated biotin defeats the purpose of using a small peptides tag, as well as it restricts the methodology to cell-surface proteins. In a separate effort, Irwin Chen in our lab found that BirA can also efficiently ligate a ketone isostere of biotin[35]. Because ketones are absent from cell surfaces, the enzymatically introduced ketone can be chemoselectively derivatized with hydrazide- or hydroxylamine probes. Unfortunately, ketones are absent from the cell-surface but are present in the cytosol (e.g., free sugars, pyruvate), which once again limits the methodology to labeling membrane proteins. Additionally, the slow kinetics of the ketone-hydrazide (or hydroxylamine) reaction results in extremely long labeling times. The transglutaminase methodology was developed by Chi-Wang Lin in our laboratory, and it uses guinea pig transglutaminase (gpTGase) to catalyze the calcium-dependent ligation of amine-containing probes to a glutamine side-chain within a seven-amino acid peptide recognition sequence[34]. Although gpTGase-mediated cell-surface labeling is fast, the labeling specificity is not very high. The lack of labeling specificity means that the methodology cannot be extended to the cytosol, where gpTGase is likely to transfer the amine-containing probe to other protein substrates. In Chapter 7, we perform a side-by-

side comparison of the gpTGase, BirA and lipoic acid ligase labeling methodologies that illustrates the main weaknesses and strengths of each one of them.

Finally, the formylglycine generating enzyme[69] and farnesyl transferase[70] have also been exploited for protein labeling, although their use has thus far only been demonstrated *in vitro*. The natural role of the formylglycine generating enzyme is to generate the sulfatase active site by converting a conserved cysteine to formylglycine. The Bertozzi lab has used a six-amino acid sulfatase motif (the "aldehyde tag") to introduce an aldehyde functionality into proteins[69]. As described above for BirA-labeling, the aldehyde moiety can then be chemoselectively derivatized with hydroxylamine- or hydrazide-probes. Farnesyl transferase appends farnesyl moieties to proteins carrying a four-amino acid farnesylation motif. Duckworth and co-workers used farnesyl transferase to label a protein C-terminally tagged with the farnesylation motif with an alkyne-containing farnesyl substrate[70]. By using the bio-orthogonal [3+2] cycloaddition reaction, the authors could then fluorescently-label the protein with an azide-derivatized Texas red fluorophore.

Although it has been shown that enzyme-mediated labeling can result in faster and specific labeling of proteins tagged with only a small peptide tag, all the currently available methodologies are, for one reason or another, restricted to the cell-surface. There is therefore a need to develop a new method that retains the positive features of the BirA and PPTase techniques, while allowing labeling inside living cells. In Chapters 6, 7, and 8 we will describe how we have accomplished this by using the enzyme lipoic acid ligase.

Other approaches to chemically modifying proteins *in vivo*

Unnatural amino acid mutagenesis[71] has also been used to label proteins in living cells with several photocrosslinkers (e.g., benzophenone, and aryl azide)[72-74] and fluorophores (e.g., dansylalanine and 7-hydroxycoumarin)[75,76]. Several reports have also combined unnatural amino acid mutagenesis with the bio-orthogonal ligations between ketones and hydrazides, and azides and alkynes[77-79]. Although unnatural amino acid

mutagenesis stands out among all other protein labeling methods because of its perfect specificity and because it uses the smallest possible tag (i.e., a single amino acid), its application is still limited in mammalian cells because of the prevalence of amber stop codons and the low suppression efficiencies, which give rise to truncated by-products[80].

Some recent efforts have adapted protein splicing for labeling proteins with small molecules. Protein splicing is a post-translational modification wherein a protein autocatalytically rearranges to excise an internal segment (an intein) and ligate the flanking N- and C-terminal sequences, or exteins[81]. Protein splicing has been used for protein labeling in two different ways. In the first approach, the N-terminal cysteine of the spliced intein is derivatized via native chemical ligation with a membrane-permeant thioester functionalized probe[82]. Although native chemical ligation displays certain chemoselectivity for N-terminal cysteines over internal cysteines of proteins, the latter can be a source of background staining. Nonetheless, Chattopadhaya and co-workers recently reported the use of native chemical ligation for labeling intracellular proteins in both, bacterial and mammalian cells with fluorescein, rhodamine, and biotin[83]. As demonstrated in the previous report, one of the main advantages of this method is the versatility for probe structure. The main drawbacks lie in the facts that, although temporary, the intein tag is large and the rates of splicing and thioester ligation are slow (i.e., it usually requires several hours).

In the second approach, developed in the Muir laboratory, site-specific protein modification is achieved via protein trans-splicing[84]. In this case, the protein of interest is fused to one half of the naturally occurring Ssp DnaE split intein, whereas the other half of the intein is conjugated to a synthetic molecule, and is introduced into the cell using a protein transduction domain. Inside the cell, the two halves of the intein reconstitute and cause protein trans-splicing, which results in ligation of the synthetic molecule to the protein of interest. The strategy has thus far only been reported once, for the incorporation of a FLAG epitope onto GFP in live Chinese hamster ovary cells[84]. The most likely reason for the lack of further demonstrations, specially using fluorescence tags, is the difficulty in washing out the excess unreacted fluorophore-intein. Additionally, the strategy suffers from the same drawbacks of using a large protein tag and of the slow kinetics of protein splicing.

Conclusions

Many protein labeling methods have been developed in recent years[14,15], but none currently allows the covalent attachment of small fluorophores of any structure onto cell surface proteins modified only by a small peptide tag, with short labeling times and with extremely high specificity over a wide range of expression levels and labeling conditions. To address this shortcoming, we developed a new protein labeling method based on the *E. coli* enzyme lipoic acid ligase (LplA)[85].

Table 5-1: Comparison of methods for genetic targeting of chemical probes in live cells.

Tag name	Ligase	Tag Size	Covalent attachment	Protein Context	Comments	Refs.
Single amino acid-directed Targeting						
Unnatural amino acid mutagenesis	NA	0	Yes	Intracelullar	Dominant negative effect due to truncations	71-80
Non-Covalent Peptide-Directed Targeting						
Tetracysteine	NA	6-10	No	Membrane	Long labeling times. Arsenic toxicity	22, 23, 39-47
Texas Red binding peptide	NA	23-38	No	Intracellular	Texas Red binds non-specifically to mitochondrial membranes	48
Hexahistidine	NA	6-12	No	Membrane* & intracellular	Ni^{2+} quenches fluorescence and it is toxic	16-18, 49, 50
Polyaspartate	NA	8-16	No	Membrane		19-21
Lanthanide binding peptide	NA	15-20	No	*In vitro*	Lanthanide toxicity. Ultraviolet excitation	51-55
Streptavidin and bungarotoxin binding peptides	NA	13	No	Membrane	Although it uses a small peptide to direct targeting, the final tag is large due to streptavidin	56
Non-Covalent Protein-Directed Targeting						
FKBP12(F36V)	NA	98	No (1)	Membrane		24

DHFR	NA	157	No	Intracellular		25, 26, 57, 58
Covalent Protein-Directed Targeting						
SNAP-tag & CLIP-Tag	O^6-alkylguanine-DNA alkyltransferase	182	Yes	Membrane & intracellular		27-29, 59
HaloTag	Dehalogenase	296	Yes	Membrane & intracellular		30
Cutinase	Cutinase	213	Yes	Membrane		61
Native chemical ligation	NA	~ 150	No	Intracellular	Chemical ligation is slow. Cannot work in oxidizing compartments	82, 83
Protein *trans* splicing	NA	~ 123	No	Intracellular	Splicing is slow. Cannot work in oxidizing compartments	84
Enzyme-mediated Peptide-Directed Targeting						
A1/S6	AcpS/Sfp PPTases	11	Yes	Membrane	Fast and specific	31-33, 62-65
Q-tag	Transglutaminase	7	Yes	Membrane	Fast but not very specific	34
AP	Biotin ligase	15	Yes	Membrane	Slow but highly specific	35-37, 66-68
LAP	Lipoic acid ligase	22	Yes	Membrane & Intracellular	Fast and specific	38
Farnesylation motif	Protein farnesyltransferase	4	Yes	*In vitro*		69
Aldehyde tag	Formylglycine generating enzyme	6	Yes	*In vitro*		70

References

1. Lippincott-Schwartz,J. & Patterson,G.H. Development and use of fluorescent protein markers in living cells. *Science* **300**, 87-91 (2003).

2. Lisenbee,C.S., Karnik,S.K. & Trelease,R.N. Overexpression and mislocalization of a tail-anchored GFP redefines the identity of peroxisomal ER. *Traffic.* **4**, 491-501 (2003).

3. Marguet,D. *et al.* Lateral diffusion of GFP-tagged H2Ld molecules and of GFP-TAP1 reports on the assembly and retention of these molecules in the endoplasmic reticulum. *Immunity.* **11**, 231-240 (1999).

4. Brock,R., Hamelers,I.H. & Jovin,T.M. Comparison of fixation protocols for adherent cultured cells applied to a GFP fusion protein of the epidermal growth factor receptor. *Cytometry* **35**, 353-362 (1999).

5. McLean,A.J. & Milligan,G. Ligand regulation of green fluorescent protein-tagged forms of the human beta(1)- and beta(2)-adrenoceptors; comparisons with the unmodified receptors. *Br. J. Pharmacol.* **130**, 1825-1832 (2000).

6. Debant,A., Ponzio,G., Clauser,E., Contreres,J.O. & Rossi,B. Receptor cross-linking restores an insulin metabolic effect altered by mutation on tyrosine 1162 and tyrosine 1163. *Biochemistry* **28**, 14-17 (1989).

7. Weiss,A. & Littman,D.R. Signal transduction by lymphocyte antigen receptors. *Cell* **76**, 263-274 (1994).

8. Anderson,R.G., Brown,M.S., Beisiegel,U. & Goldstein,J.L. Surface distribution and recycling of the low density lipoprotein receptor as visualized with antireceptor antibodies. *J. Cell Biol.* **93**, 523-531 (1982).

9. Neri,D. *et al.* Targeting by affinity-matured recombinant antibody fragments of an angiogenesis associated fibronectin isoform. *Nat. Biotechnol.* **15**, 1271-1275 (1997).

10. Rader,C. *et al.* The rabbit antibody repertoire as a novel source for the generation of therapeutic human antibodies. *J. Biol. Chem.* **275**, 13668-13676 (2000).

11. Lidke,D.S. *et al.* Quantum dot ligands provide new insights into erbB/HER receptor-mediated signal transduction. *Nat. Biotechnol.* **22**, 198-203 (2004).

12. Ehrlich,M. *et al.* Endocytosis by random initiation and stabilization of clathrin-coated pits. *Cell* **118**, 591-605 (2004).

13. Barak,L.S. & Webb,W.W. Fluorescent low density lipoprotein for observation of dynamics of individual receptor complexes on cultured human fibroblasts. *J. Cell Biol.* **90**, 595-604 (1981).

14. Marks,K.M. & Nolan,G.P. Chemical labeling strategies for cell biology. *Nat. Methods* **3**, 591-596 (2006).

15. O'hare,H.M., Johnsson,K. & Gautier,A. Chemical probes shed light on protein function. *Curr. Opin. Struct. Biol.* (2007).

16. Guignet,E.G., Hovius,R. & Vogel,H. Reversible site-selective labeling of membrane proteins in live cells. *Nat. Biotechnol.* **22**, 440-444 (2004).

17. Guignet,E.G., Segura,J.M., Hovius,R. & Vogel,H. Repetitive reversible labeling of proteins at polyhistidine sequences for single-molecule imaging in live cells. *Chemphyschem.* **8**, 1221-1227 (2007).

18. Lata,S., Gavutis,M., Tampe,R. & Piehler,J. Specific and stable fluorescence labeling of histidine-tagged proteins for dissecting multi-protein complex formation. *J. Am. Chem. Soc.* **128**, 2365-2372 (2006).

19. Honda,K., Nakata,E., Ojida,A. & Hamachi,I. Ratiometric fluorescence detection of a tag fused protein using the dual-emission artificial molecular probe. *Chem. Commun. (Camb.)* 4024-4026 (2006).

20. Nonaka,H., Tsukiji,S., Ojida,A. & Hamachi,I. Non-enzymatic covalent protein labeling using a reactive tag. *J. Am. Chem. Soc.* **129**, 15777-15779 (2007).

21. Ojida,A. *et al.* Oligo-Asp tag/Zn(II) complex probe as a new pair for labeling and fluorescence imaging of proteins. *J. Am. Chem. Soc.* **128**, 10452-10459 (2006).

22. Gaietta,G. *et al.* Multicolor and electron microscopic imaging of connexin trafficking. *Science* **296**, 503-507 (2002).

23. Griffin,B.A., Adams,S.R. & Tsien,R.Y. Specific covalent labeling of recombinant protein molecules inside live cells. *Science* **281**, 269-272 (1998).

24. Marks,K.M., Braun,P.D. & Nolan,G.P. A general approach for chemical labeling and rapid, spatially controlled protein inactivation. *Proc. Natl. Acad. Sci. U. S. A* **101**, 9982-9987 (2004).

25. Miller,L.W., Sable,J., Goelet,P., Sheetz,M.P. & Cornish,V.W. Methotrexate conjugates: a molecular in vivo protein tag. *Angew. Chem. Int. Ed Engl.* **43**, 1672-1675 (2004).

26. Miller,L.W., Cai,Y., Sheetz,M.P. & Cornish,V.W. In vivo protein labeling with trimethoprim conjugates: a flexible chemical tag. *Nat. Methods* **2**, 255-257 (2005).

27. Gautier,A. *et al.* An engineered protein tag for multiprotein labeling in living cells. *Chem. Biol.* **15**, 128-136 (2008).

28. Keppler,A. *et al.* A general method for the covalent labeling of fusion proteins with small molecules in vivo. *Nat. Biotechnol.* **21**, 86-89 (2003).

29. Keppler,A., Pick,H., Arrivoli,C., Vogel,H. & Johnsson,K. Labeling of fusion proteins with synthetic fluorophores in live cells. *Proc. Natl. Acad. Sci. U. S. A* **101**, 9955-9959 (2004).

30. Los,G.V. & Wood,K. The HaloTag: a novel technology for cell imaging and protein analysis. *Methods Mol. Biol.* **356**, 195-208 (2007).

31. George,N., Pick,H., Vogel,H., Johnsson,N. & Johnsson,K. Specific labeling of cell surface proteins with chemically diverse compounds. *J. Am. Chem. Soc.* **126**, 8896-8897 (2004).

32. Yin,J., Liu,F., Li,X. & Walsh,C.T. Labeling proteins with small molecules by site-specific posttranslational modification. *J. Am. Chem. Soc.* **126**, 7754-7755 (2004).

33. Zhou,Z. *et al.* Genetically encoded short peptide tags for orthogonal protein labeling by Sfp and AcpS phosphopantetheinyl transferases. *ACS Chem. Biol.* **2**, 337-346 (2007).

34. Lin,C.W. & Ting,A.Y. Transglutaminase-catalyzed site-specific conjugation of small-molecule probes to proteins in vitro and on the surface of living cells. *J. Am. Chem. Soc.* **128**, 4542-4543 (2006).

35. Chen,I., Howarth,M., Lin,W. & Ting,A.Y. Site-specific labeling of cell surface proteins with biophysical probes using biotin ligase. *Nat. Methods* **2**, 99-104 (2005).

36. Howarth,M., Takao,K., Hayashi,Y. & Ting,A.Y. Targeting quantum dots to surface proteins in living cells with biotin ligase. *Proc. Natl. Acad. Sci. U. S. A* **102**, 7583-7588 (2005).

37. Howarth,M. *et al.* A monovalent streptavidin with a single femtomolar biotin binding site. *Nature Methods* **3**, 267-273 (2006).

38. Fernandez-Suarez,M. *et al.* Redirecting lipoic acid ligase for cell surface protein labeling with small-molecule probes. *Nat. Biotechnol.* **25**, 1483-1487 (2007).

39. Griffin,B.A., Adams,S.R. & Tsien,R.Y. Specific covalent labeling of recombinant protein molecules inside live cells. *Science* **281**, 269-272 (1998).

40. Adams,S.R. *et al.* New biarsenical ligands and tetracysteine motifs for protein labeling in vitro and in vivo: synthesis and biological applications. *J. Am. Chem. Soc.* **124**, 6063-6076 (2002).

41. Tour,O., Meijer,R.M., Zacharias,D.A., Adams,S.R. & Tsien,R.Y. Genetically targeted chromophore-assisted light inactivation. *Nat. Biotechnol.* **21**, 1505-1508 (2003).

42. Ju,W. *et al.* Activity-dependent regulation of dendritic synthesis and trafficking of AMPA receptors. *Nat. Neurosci.* **7**, 244-253 (2004).

43. Panchal,R.G. *et al.* In vivo oligomerization and raft localization of Ebola virus protein VP40 during vesicular budding. *Proc. Natl. Acad. Sci. U. S. A* **100**, 15936-15941 (2003).

44. Poskanzer,K.E., Marek,K.W., Sweeney,S.T. & Davis,G.W. Synaptotagmin I is necessary for compensatory synaptic vesicle endocytosis in vivo. *Nature* **426**, 559-563 (2003).

45. Turville,S.G., Aravantinou,M., Stossel,H., Romani,N. & Robbiani,M. Resolution of de novo HIV production and trafficking in immature dendritic cells. *Nat. Methods* **5**, 75-85 (2008).

46. Stroffekova,K., Proenza,C. & Beam,K.G. The protein-labeling reagent FLASH-EDT2 binds not only to CCXXCC motifs but also non-specifically to endogenous cysteine-rich proteins. *Pflugers Arch.* **442**, 859-866 (2001).

47. Martin,B.R., Giepmans,B.N., Adams,S.R. & Tsien,R.Y. Mammalian cell-based optimization of the biarsenical-binding tetracysteine motif for improved fluorescence and affinity. *Nat. Biotechnol.* **23**, 1308-1314 (2005).

48. Marks,K.M., Rosinov,M. & Nolan,G.P. In vivo targeting of organic calcium sensors via genetically selected peptides. *Chem. Biol.* **11**, 347-356 (2004).

49. Goldsmith,C.R., Jaworski,J., Sheng,M. & Lippard,S.J. Selective labeling of extracellular proteins containing polyhistidine sequences by a fluorescein-nitrilotriacetic acid conjugate. *J. Am. Chem. Soc.* **128**, 418-419 (2006).

50. Hauser,C.T. & Tsien,R.Y. A hexahistidine-Zn2+-dye label reveals STIM1 surface exposure. *Proc. Natl. Acad. Sci. U. S. A* **104**, 3693-3697 (2007).

51. Franz,K.J., Nitz,M. & Imperiali,B. Lanthanide-binding tags as versatile protein coexpression probes. *Chembiochem.* **4**, 265-271 (2003).

52. Martin,L.J. *et al.* Double-lanthanide-binding tags: design, photophysical properties, and NMR applications. *J. Am. Chem. Soc.* **129**, 7106-7113 (2007).

53. Silvaggi,N.R., Martin,L.J., Schwalbe,H., Imperiali,B. & Allen,K.N. Double-lanthanide-binding tags for macromolecular crystallographic structure determination. *J. Am. Chem. Soc.* **129**, 7114-7120 (2007).

54. Sculimbrene,B.R. & Imperiali,B. Lanthanide-binding tags as luminescent probes for studying protein interactions. *J. Am. Chem. Soc.* **128**, 7346-7352 (2006).

55. Reynolds,A.M., Sculimbrene,B.R. & Imperiali,B. Lanthanide-binding tags with unnatural amino acids: sensitizing Tb3+ and Eu3+ luminescence at longer wavelengths. *Bioconjug. Chem.* **19**, 588-591 (2008).

56. McCann,C.M., Bareyre,F.M., Lichtman,J.W. & Sanes,J.R. Peptide tags for labeling membrane proteins in live cells with multiple fluorophores. *Biotechniques* **38**, 945-952 (2005).

57. Sasso,S.P., Gilli,R.M., Sari,J.C., Rimet,O.S. & Briand,C.M. Thermodynamic study of dihydrofolate reductase inhibitor selectivity. *Biochim. Biophys. Acta* **1207**, 74-79 (1994).

58. Calloway,N.T. *et al.* Optimized fluorescent trimethoprim derivatives for in vivo protein labeling. *Chembiochem.* **8**, 767-774 (2007).

59. Gronemeyer,T., Chidley,C., Juillerat,A., Heinis,C. & Johnsson,K. Directed evolution of O6-alkylguanine-DNA alkyltransferase for applications in protein labeling. *Protein Eng Des Sel* **19**, 309-316 (2006).

60. Zhang,Y. *et al.* HaloTag protein-mediated site-specific conjugation of bioluminescent proteins to quantum dots. *Angew. Chem. Int. Ed Engl.* **45**, 4936-4940 (2006).

61. Bonasio,R. *et al.* Specific and covalent labeling of a membrane protein with organic fluorochromes and quantum dots. *Proc. Natl. Acad. Sci. U. S. A* **104**, 14753-14758 (2007).

62. Meier,J.L., Mercer,A.C., Rivera,H., Jr. & Burkart,M.D. Synthesis and evaluation of bioorthogonal pantetheine analogues for in vivo protein modification. *J. Am. Chem. Soc.* **128**, 12174-12184 (2006).

63. Vivero-Pol,L., George,N., Krumm,H., Johnsson,K. & Johnsson,N. Multicolor imaging of cell surface proteins. *J. Am. Chem. Soc.* **127**, 12770-12771 (2005).

64. Yin,J. *et al.* Single-cell FRET imaging of transferrin receptor trafficking dynamics by Sfp-catalyzed, site-specific protein labeling. *Chem. Biol.* **12**, 999-1006 (2005).

65. Jacquier,V., Prummer,M., Segura,J.M., Pick,H. & Vogel,H. Visualizing odorant receptor trafficking in living cells down to the single-molecule level. *Proc. Natl. Acad. Sci. U. S. A* **103**, 14325-14330 (2006).

66. Howarth,M. & Ting,A.Y. Imaging proteins in live mammalian cells with biotin ligase and monovalent streptavidin. *Nat. Protoc.* **3**, 534-545 (2008).

67. Howarth,M. *et al.* Monovalent, reduced-size quantum dots for imaging receptors on living cells. *Nat. Methods* **5**, 397-399 (2008).

68. Chen,I., Choi,Y.A. & Ting,A.Y. Phage display evolution of a peptide substrate for yeast biotin ligase and application to two-color quantum dot labeling of cell surface proteins. *J. Am. Chem. Soc.* **129**, 6619-6625 (2007).

69. Carrico,I.S., Carlson,B.L. & Bertozzi,C.R. Introducing genetically encoded aldehydes into proteins. *Nat. Chem. Biol.* **3**, 321-322 (2007).

70. Duckworth,B.P., Zhang,Z., Hosokawa,A. & Distefano,M.D. Selective labeling of proteins by using protein farnesyltransferase. *Chembiochem.* **8**, 98-105 (2007).

71. Wang,L., Xie,J. & Schultz,P.G. Expanding the genetic code. *Annu. Rev. Biophys. Biomol. Struct.* **35**, 225-249 (2006).

72. Chin,J.W. & Schultz,P.G. In vivo photocrosslinking with unnatural amino Acid mutagenesis. *Chembiochem.* **3**, 1135-1137 (2002).

73. Farrell,I.S., Toroney,R., Hazen,J.L., Mehl,R.A. & Chin,J.W. Photo-cross-linking interacting proteins with a genetically encoded benzophenone. *Nat. Methods* **2**, 377-384 (2005).

74. Hino,N. *et al.* Protein photo-cross-linking in mammalian cells by site-specific incorporation of a photoreactive amino acid. *Nat. Methods* **2**, 201-206 (2005).

75. Summerer,D. *et al.* A genetically encoded fluorescent amino acid. *Proc. Natl. Acad. Sci. U. S. A* **103**, 9785-9789 (2006).

76. Wang,J., Xie,J. & Schultz,P.G. A genetically encoded fluorescent amino acid. *J. Am. Chem. Soc.* **128**, 8738-8739 (2006).

77. Chin,J.W. *et al.* Addition of p-azido-L-phenylalanine to the genetic code of Escherichia coli. *J. Am. Chem. Soc.* **124**, 9026-9027 (2002).

78. Deiters,A. *et al.* Adding amino acids with novel reactivity to the genetic code of Saccharomyces cerevisiae. *J. Am. Chem. Soc.* **125**, 11782-11783 (2003).

79. Zhang,Z. *et al.* A new strategy for the site-specific modification of proteins in vivo. *Biochemistry* **42**, 6735-6746 (2003).

80. Liu,W., Brock,A., Chen,S., Chen,S. & Schultz,P.G. Genetic incorporation of unnatural amino acids into proteins in mammalian cells. *Nat. Methods* **4**, 239-244 (2007).

81. Noren,C.J., Wang,J. & Perler,F.B. Dissecting the Chemistry of Protein Splicing and Its Applications. *Angew. Chem. Int. Ed Engl.* **39**, 450-466 (2000).

82. Yeo,D.S. *et al.* Cell-permeable small molecule probes for site-specific labeling of proteins. *Chem. Commun. (Camb.)* 2870-2871 (2003).

83. Chattopadhaya,S., Srinivasan,R., Yeo,D.S., Chen,G.Y. & Yao,S.Q. Site-specific covalent labeling of proteins inside live cells using small molecule probes. *Bioorg. Med. Chem.* (2008).

84. Giriat,I. & Muir,T.W. Protein semi-synthesis in living cells. *J. Am. Chem. Soc.* **125**, 7180-7181 (2003).

85. Green,D.E., Morris,T.W., Green,J., Cronan,J.E., Jr. & Guest,J.R. Purification and properties of the lipoate protein ligase of Escherichia coli. *Biochem. J.* **309 (Pt 3)**, 853-862 (1995).

Chapter 6 : Re-directing *E. coli* lipoic acid ligase for site-specific protein labeling

The work discussed in this chapter has been published in part in: M. Fernández-Suárez, H. Baruah, L. Martínez-Hernández, K. T. Xie, J. M. Baskin, C. R. Bertozzi, and A. Y. Ting. Redirecting lipoic acid ligase for cell surface protein labeling with small-molecule probes. *Nat. Biotechnol.* **25**(12), 1483-1487 (2007). The syntheses of the alkyl azide and alkyl alkyne probes were done by Laura Martínez-Hernández. Determination of the kinetics of LplA-catalyzed ligation of azide **7** to E2p and LAP was performed in collaboration with Dr. Hemanta Baruah.

Introduction

E. coli lipoic acid ligase

In the introductory Chapter 5, we described the various approaches developed in recent years to site-specifically label proteins with chemical probes in living cells. The most promising of these strategies employs an enzyme to catalyze the covalent ligation of the chemical probe to a peptide recognition sequence fused to the protein of interest. Our lab has previously developed protein labeling methodologies based on the biotin ligase and transglutaminase enzymes, but both methodologies are restricted to labeling cell-surface proteins. In this chapter, we describe how we have re-directed the *E. coli* enzyme lipoic acid ligase (LplA) for the purpose of protein labeling. In principle, LplA should retain the excellent labeling specificity of biotin ligase, with the additional benefit that it has the potential to work inside cells.

In *E. coli*, LplA catalyzes the ATP-dependent covalent ligation of lipoic acid to one of three proteins substrates[1]. The mechanism of lipoic acid transfer proceeds in two distinct steps (Figure 6-1). First, the enzyme activates the carboxylic acid of the lipoic acid cofactor using ATP and releasing pyrophosphate (PPi) in the process. In the second step, the reactive adenylate ester intermediate (lipoyl-AMP) is attacked by the ε-amino group of a lysine residue in the lipoate-acceptor protein (e.g., E2p) to effect lipoic acid transfer.

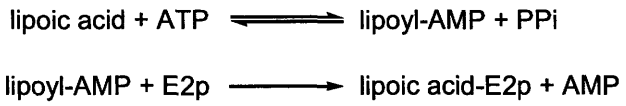


Figure 6-1: Site-specific lipoic acid transfer catalyzed by *E. coli* lipoic acid ligase. The mechanism of lipoylation occurs in two steps. In the first step, the carboxylic acid of lipoic acid is activated by reaction with ATP to form the lipoic acid adenylate ester (lipoyl-AMP). In the next step, the reactive lysine of the substrate (e.g., the E2p protein) attacks lipoyl-AMP, effecting lipoic acid transfer.

R-(+)-lipoic acid is a cofactor widely distributed among living organisms. In *E. coli*, lipoic acid is covalently attached to the E2 dihydrolipoamide acetyltransferase subunit (E2p) of the pyruvate dehydrogenase complex (PDH), to the E2 dihydrolipoamide succinyltransferase subunit (E2o) of the 2-oxoglutarate dehydrogenase

complex (2OGDH), and to the H-protein of the glycine cleavage system[2]. Additionally, in mammals, lipoic acid is also a cofactor of the E2 acyltransferase subunit of the branched-chain α-ketoacid dehydrogenase complex (E2b)[3]. The four enzyme complexes are involved in oxidative metabolism and, in mammals, located in the mitochondria. In the PDH and 2OGDH complexes, the E1 subunits catalyze the initial decarboxylation of the 2-oxo acid (pyruvate or 2-oxoglutarate), using thiamine diphosphate (TPP) as a cofactor, and then reductively acylate the lipoyl group attached to a lysine in the E2 subunit[4]. E2 is then responsible for transferring the acyl group on to coenzyme A (CoA), to form acetyl-CoA or succinyl-CoA. The dihydrolipoyl group left on E2 is finally reoxidized to the dithiolane ring by the flavoprotein E3, with the nicotinamide adenine dinucleotide (NAD^+) as the ultimate electron acceptor. The lipoyllysine arm attached to the E2 subunits plays a pivotal role in the reaction sequence in the complexes, shuttling the reaction intermediate and reducing equivalents among the active sites of the subunits.

There are two pathways involved in protein lipoylation. The first pathway uses exogenously supplied lipoic acid. This is the pathway described above, which, in *E. coli*, is catalyzed by LplA. In contrast to LplA, which can catalyze both, the adenylation of lipoic acid and its transfer to the protein substrate, mammalian lipoyltransferases (or lipoate protein ligases, LPLs) can only catalyze the second step[3,5]. Therefore, the contribution of another enzyme, the lipoate-activating enzyme, is required for protein lipoylation[6]. In *E. coli*, a second lipoylation pathway uses endogenously synthesized lipoic acid[7]. Lipoic acid is synthesized from the octanoyl-acyl carrier protein provided by the fatty acid synthesis pathway. The octanoyl moiety is first transferred to the lipoate-acceptor protein by the action of the LipB enzyme[8]. Then, the sulfur atoms are inserted at the C6 and C8 positions of the octanoyl moiety by LipA, an *S*-adenosyl methionine-dependent [Fe-S] cluster containing enzyme[9]. Although not yet demonstrated, it has been suggested that this second lipoylation pathway is also present in mammals[10].

Three dimensional structures of LPLs from various species have been solved[11-14]. Bacterial LplAs are formed by two domains, a large N-terminal domain and a small C-terminal domain. This is the case for example of *Escherichia coli* (1X2G)[11] and *Streptococcus pneumoniae* (1VQZ). LplAs from Archaea, like *Thermoplasma acidophilum* (Ta) (2ART and 2C8M)[13,14] and *Sulfolobus solfataricus* are however

composed by just one domain, corresponding to the bacterial N-terminal domain. This lack of a C-terminal domain may explain the observation that Ta LplA can only catalyze the first step of lipoic acid adenylation[14]. In Ta, a second protein has been suggested to perform the transfer of the adenylate intermediate[14]. The lipoic acid binding site and the catalytic center are situated in the N-terminal domain in all LplAs. The case of the mammalian LPLs is however harder to explain based on the enzyme structures. Although the mammalian lipoyltransferases share about 30% identity with that of *E. coli* LplA[12], showing specially high identity in the N-terminal half, mammalian LPLs have been suggested to be unable to activate lipoic acid[5,6]. It remains to be shown how this lack of function relates to their structure.

Using LplA for site-specific protein labeling

E. coli LplA thus catalyzes the type of reaction that is needed for protein labeling: ligation of a chemical probe onto a protein substrate. However, to be useful for labeling proteins in living cells, LplA still needed to meet three requirements:

1. Because lipoic acid is not amenable for imaging, we needed to find a small-molecule substrate, such as a fluorophore or a functional group handle (e.g., azide, ketones, alkynes), that can be ligated by LplA and that is useful for imaging or protein-protein interaction studies.
2. LplA naturally catalyzes the transfer of lipoic acid to one of three protein substrates. In order to minimize perturbation of the target protein, we needed to find a small peptide substrate that can be recognized by LplA and be fused to the target protein with minimal steric interference.
3. LplA should also display excellent specificity towards its peptide substrate, such that it does not recognize any endogenous protein within the mammalian proteome.

In the previous section, we have described how LplA naturally exhibits extremely high sequence specificity towards its protein substrate, as it only transfers lipoic acid to three proteins in *E. coli*. Thus, we expected LplA to easily comply with the third requirement. In contrast, and to our advantage, previous work had shown that the enzyme

has a much more relaxed specificity towards its small-molecule substrate. In *E. coli*, LplA can also ligate octanoate and selenolipoate, if the bacteria are grown in the absence of its natural substrate R-(+)-lipoic acid[15]. Furthermore, the enzyme has been shown to accept other analogs of lipoic acid *in vitro*. Green and co-workers showed that LplA can ligate, in order of highest to lowest ligation efficiency, 6-thiooctanoate, S-(−)-lipoate, dihydrolipoate, 8-methyl-lipoate, octanoate, and selenolipoate (Figure 6-2A)[1]. These data suggested that the small-molecule binding site in LplA has considerable plasticity, and encouraged us to look for other small-molecule probes that could be incorporated by the enzyme.

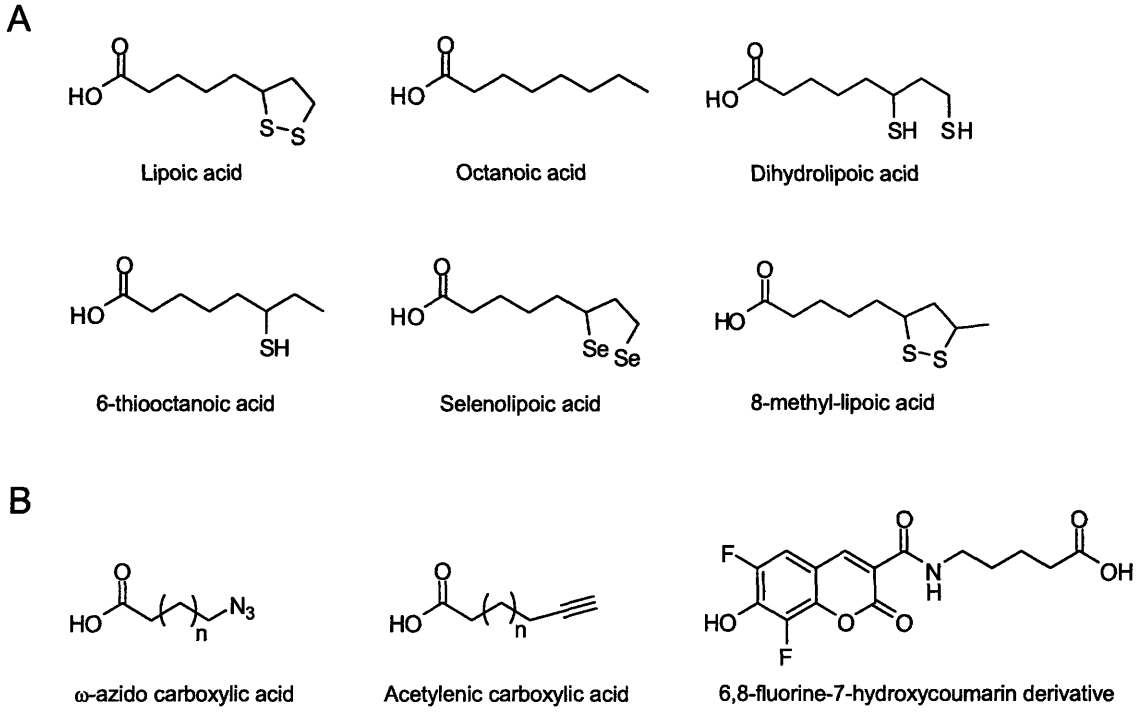


Figure 6-2: Structure of lipoic acid and several lipoic acid analogs. (**A**) Structure of lipoic acid and analogs reported to be ligated by LplA *in vitro*. (**B**) Structure of small-probes useful for protein labeling that have been tested in our lab for incorporation by LplA.

We considered a range of small molecule structures to replace lipoic acid. Direct ligation of a fluorophore, such as the coumarin probe depicted in Figure 6-2B, would offer a simpler and shorter labeling procedure, but incorporation of a "functional group handle" is more feasible due to the small size of the lipoate binding pocket, and provides greater versatility for subsequent incorporation of probes of any structure. Many functional group handles have been used in chemical biology, including ketones, organic

azides, and alkynes[16]. In previous work in our lab, Irwin Chen had found that biotin ligase could accept a ketone isostere of biotin, which could then be chemoselectively derivatized with hydrazide- or hydroxylamine-conjugated probes[17]. Although Irwin demonstrated that this two-step procedure could afford labeling of membrane proteins, the method is restricted to the cell-surface because ketones are absent from the cell-surface but exist naturally inside cells.

Alkyl azides and alkynes (Figure 6-2B) are useful functional groups for protein labeling because they are both absent from biological organisms, but possess selective reactivity towards each other[16]. The classic [3+2] dipolar cycloaddition reaction between azides and alkynes was developed by Huisgen and requires high temperatures or pressures to proceed efficiently[18]. Sharpless, Finn, and coworkers developed a Cu(I)-catalyzed version[19] (Figure 6-3, above), which occurs with high efficiency and selectivity in pH 7.4 water at room temperature and 1 atmosphere, even in complex environments such as whole-cell lysates, or the surface of mammalian cells, bacteria, and virus particles[19-22]. The reaction is highly toxic to cells, however, due to the requirement for Cu(I). For this reason, the Staudinger ligation between azides and triarylphosphines[23] (Figure 6-3, middle) has often been preferred for protein labeling inside mammalian cells[24-26]. Unfortunately, the second-order rate of the Staudinger ligation is ~ 30-fold lower than that of the Cu(I)-mediated cycloaddition[27], and triarylphosphines tend to rapidly oxidize in the eukaryotic cytosol when they are conjugated to fluorophores (Bertozzi, C.R. and Baskin, J.M., personal communication). To overcome these limitations, the Bertozzi lab recently developed a copper-free version of the [3+2] cycloaddition that makes use of a highly strained alkyne presented in the context of a cyclooctyne ring[28] (Figure 6-3, below). The greatly increased reactivity of the alkyne makes it undergo rapid cycloaddition with azides, even in the absence of Cu(I). Addition of one or two electron-withdrawing fluorines adjacent to the alkyne increases the speed of the ligation yet further (~ 2-fold or ~ 22-fold, respectively)[27,29,30]. In Chapter 7, we describe how we have employed this improved strain-promoted [3+2] cycloaddition reaction for cell-surface protein labeling.

In the results section of this chapter, we describe our three-stage process of LplA engineering. First, we show that wild-type LplA can efficiently ligate an alkyl azide with

similar kinetics to lipoic acid. Second, we describe the rational design of a 22-amino acid peptide substrate that is modified by LplA only at a 2-fold slower rate than the full-length protein substrate. Finally, we demonstrate that LplA does not recognize any endogenous mammalian protein.

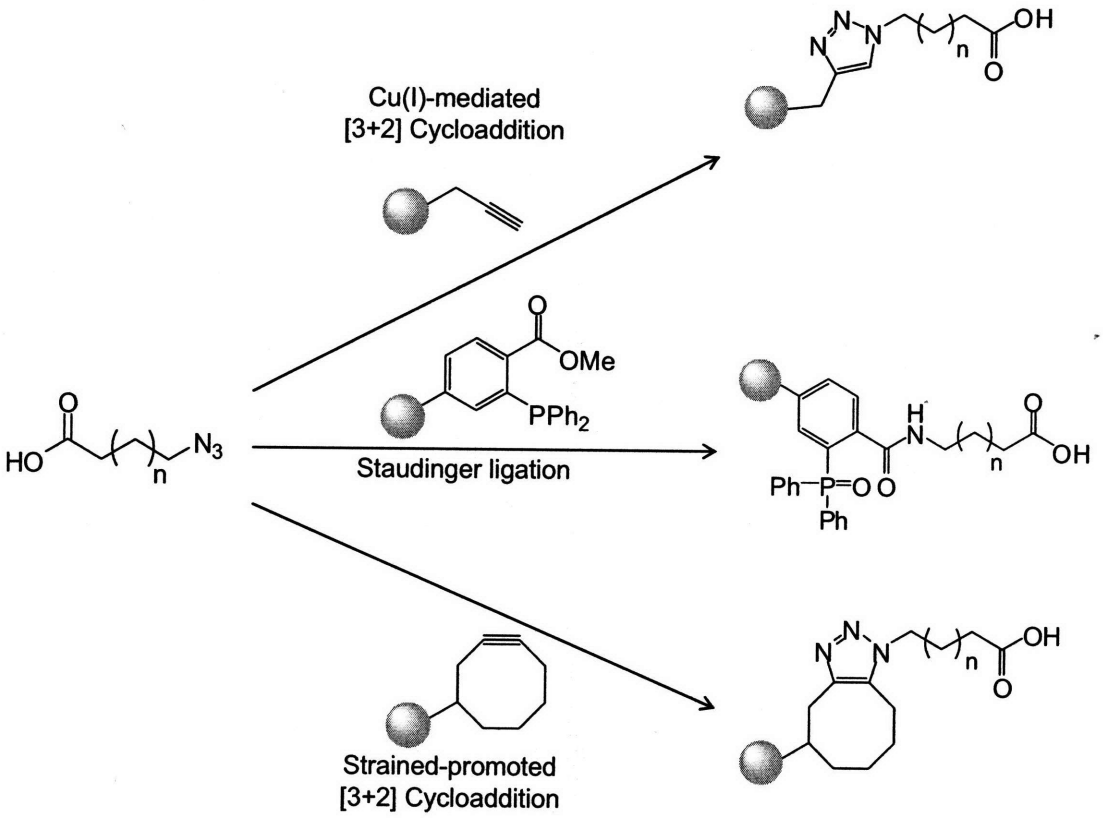


Figure 6-3: Bio-orthogonal reactions of azides. Azides react with terminal alkynes to form stable triazols. In the Sharpless version of the [3+2] cycloaddition (above), Cu(I) is added as a catalyst such that the reaction proceeds at physiological conditions. Azides can also undergo Staudinger ligation with triarylphosphines to form a stable adduct that incorporates the phosphine oxide byproduct (middle). Bertozzi's Huisgen ligation between an azide and a strained cyclooctyne proceeds at physiological conditions in the absence of copper owing to the increased reactivity of the alkyne (below).

Results

E. coli LplA accepts alkyl azide and alkyne substrates in place of lipoic acid

For the first stage of LplA engineering, we synthesized eight alkyl azide and alkyne substrates for LplA of varying lengths. The synthesis of all these probes was performed by Laura Martínez-Hernández according to the reaction scheme in Figure 6-4. The alkyl azides were synthesized either by nucleophilic substitution of the corresponding bromoalkanoic acid with sodium azide[31], or by copper-catalyzed diazo transfer onto the amine precursor[32]. The alkynes were synthesized from the bromoalkanoic acid precursor, by displacement with lithium acetylide[33].

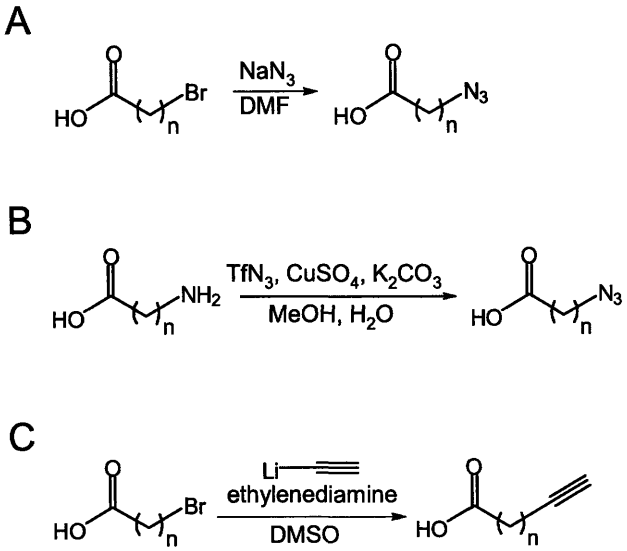
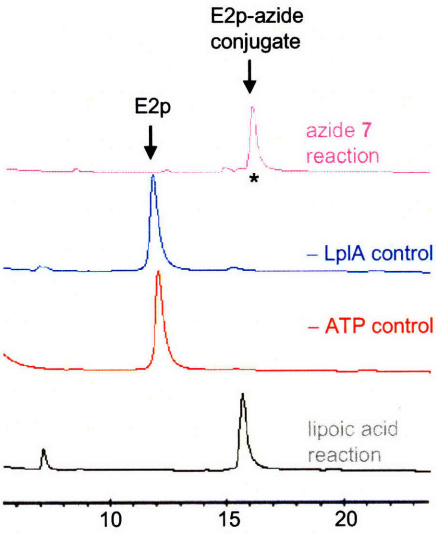


Figure 6-4: Syntheses of ω-azido and acetylenic carboxylic acid analogs of lipoic acid. (A) Alkyl azides can be prepared by nucleophilic substitution of the corresponding bromoalkanoic precursor with sodium azide. Alkyl azides with n=5 and 7–10 were prepared following this procedure. **(B)** An alkyl azide with n=6 was obtained by copper-catalyzed diazo transfer of its amine precursor. **(C)** Acetylenic carboxylic acids with n=5 and 7 were prepared from the bromoalkanoic precursor by displacement with lithium acetylide.

In order to test if these substrate analogs could be used by LplA, we developed an assay for incorporation onto a hybrid lipoyl domain derived from the full-length E2p protein[34]. The E2p subunit of the pyruvate dehydrogenase complex contains three N-terminal highly conserved tandemly repeated lipoyl domains, each containing one lysine

site of lipoylation[35]. Miles and Guest initially demonstrated that the number of lipoyl domains in E2p can be reduced from three to one without impairing catalytic activity[36]. They next cloned a hybrid lipoyl domain comprised by residues 1–33 and 238–289 from the first and third lipoyl domains, and showed that the domain was lipoylated in *E. coli* to levels comparable to those of the full-length subunit[36]. The 9 kD hybrid lipoyl domain can be recombinantly expressed in *E. coli* in high quantities[34]. We used this hybrid lipoyl domain to test incorporation of lipoic acid analogs using an HPLC assay. From hereon, we will refer to this hybrid domain as E2p domain, or simply E2p.

A



B

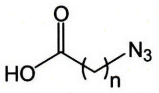
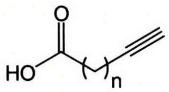
Substrate		% Conversion
lipoic acid		100 ± 3
azide 	n=5	14.9 ± 0.6
	n=6	11.3 ± 0.4
	n=7	55.4 ± 5.7
	n=8	10.2 ± 0.4
	n=9	4.57 ± 0.11
	n=10	3.28 ± 0.17
alkyne 	n=4	3.59 ± 0.23
	n=5	3.15 ± 0.19
	n=7	6.37 ± 0.15
	n=8	5.15 ± 0.23

Figure 6-5: Incorporation of lipoic acid analogs by wild-type LplA. (**A**) HPLC assay showing the ligation of lipoic acid and the azide **7** substrate to the E2p domain. The starred peak was analyzed by mass-spectrometry in Figure 6-6. (**B**) Comparison of alkyl azide and alkyne substrates of LplA. The extent of LplA-catalyzed E2p conversion in the presence of each substrate was assessed by HPLC. Conversions are given relative to lipoic acid, which is normalized to 100%.

Figure 6-5A shows that the E2p domain had a retention time of ~ 12 minutes, which increased to ~ 15-18 minutes after covalent conjugation to lipoic acid or an analog. Controls with enzyme or ATP omitted demonstrated that the product peak resulted from enzymatic ligation. We were surprised to find that all the tested alkyl azides and alkynes were incorporated to some degree by wild-type LplA (Figure 6-5B). It is worth noting that the azides were in general incorporated at a higher rate than the alkynes. Among the

alkyl azides, we also observed a clear length dependence, with the n=7 azide (azide **7**) giving the best kinetics; both shorter and longer azides were not incorporated as efficiently, perhaps because of non-optimal binding in the lipoate binding pocket.

So that we could make use of the copper-free [3+2] cycloaddition reaction with cyclooctyne probes, we selected the azide **7** probe for further characterization. Although we have not demonstrated it yet, LplA-catalyzed site-specific alkyne ligation could be useful for *in vitro* or lysate experiments, where the use of a Cu(I) catalyst accelerates the [3+2] reaction without posing a toxicity problem. We also tested other small-molecules for incorporation by wild-type LplA such as biotin, biotinamido caproate, and glutamate, but none of them was ligated by the wild-type enzyme to a detectable level.

Characterization of azide 7 ligation by LplA

To demonstrate the site-specificity of azide **7** ligation, we collected the product peak (starred) from the top HPLC trace in Figure 6-5A and we analyzed it by electrospray ionization mass-spectrometry (ESI-MS).

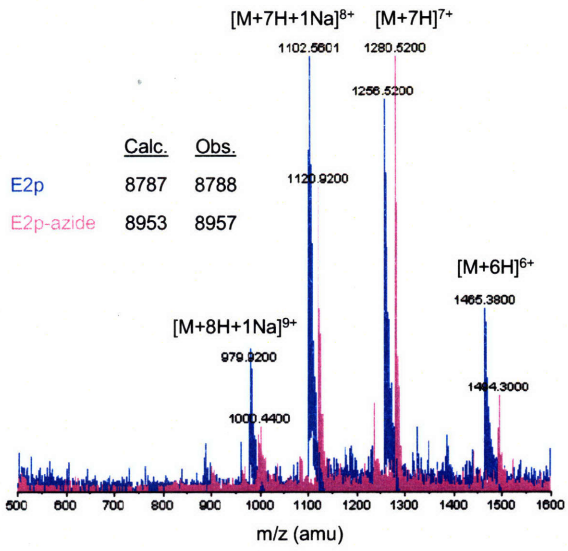


Figure 6-6: Mass-spectrometry characterization of the E2p-azide 7 conjugate. Overlay of the ESI-MS spectra of the apo-E2p (blue) and the product peak from an LplA ligation of azide **7** (magenta, from the HPLC trace in Figure 6-5). Peaks for the +6, +7, +8, and +9 charge states are observed. The observed mass for the apo-protein matched the calculated mass. The observed mass for the modified protein corresponded to the addition of a single molecule of azide **7** with the concomitant loss of water.

Figure 6-6 shows that the observed mass for the E2p-azide **7** product corresponded to the addition of a single molecule of azide **7** with the concomitant loss of water. This assay demonstrated that the ligation of azide **7** is site-specific since no ligation was observed to any of the other lysines present in the E2p domain. The spectrum for the unmodified domain is also shown for comparison. The observed mass for the apo protein matched the calculated mass.

We next wished to determine the kinetics of the LplA-catalyzed ligation of azide **7** to E2p. We first confirmed the suitability of the HPLC assay to perform kinetic analysis by determining the initial rate of the lipoic acid ligation, which we could compare with previously reported values (Figure 6-7). We obtained an initial rate of $0.25 \pm 0.3 \times 10^{-2}$ s^{-1}, which was in accordance to the reported value of 0.5 s^{-1} (Ref. 1), thus validating our assay.

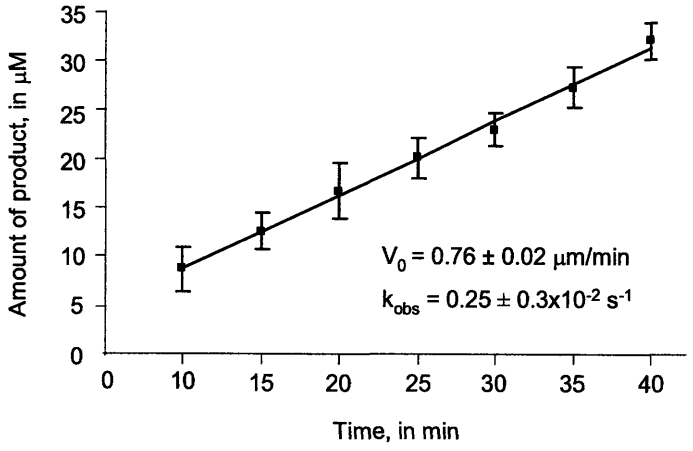


Figure 6-7: Time course of lipoic acid ligation to E2p using 50 nM LplA. Reactions were incubated at 30 °C under the following reaction conditions: 200 μM E2p, 50 nM LplA, 750 μM lipoic acid, 1 mM ATP, and 2 mM magnesium acetate in 25 mM sodium phosphate pH 7.0. Aliquots were removed at the indicated times, quenched with EDTA, and analyzed by HPLC. Each data point represents the average of three experiments. Error bars, 1 s.d.

To determine the k_{cat} and K_m for the ligation of azide **7**, we measured the initial reaction rates at a range of azide **7** concentrations (25–750 μM). From the Michaelis-Menten plot (Figure 6-8) we determined a V_{max} of 1.33 ± 0.04 μM/min using 200 nM LplA, corresponding to a k_{cat} of $0.11 \pm 0.03 \times 10^{-1}$ s^{-1}. This $k_{cat}$was only 2.3-fold lower than the measured initial rate for lipoic acid ligation. The K_m for azide **7** was 127 ± 11 μM, which is 75- or 30-fold higher than the two reported K_m values for lipoic acid (1.7

μM[1] and 4.5 μM[11]). Thus, replacement of lipoic acid with the linear azide **7** substrate does not significantly change k_{cat}, although K_m increases by one to two orders of magnitude. For labeling reactions, however, it is feasible to use azide **7** at concentrations higher than the measured K_m, so that LplA-catalyzed ligation occurs at maximal velocity.

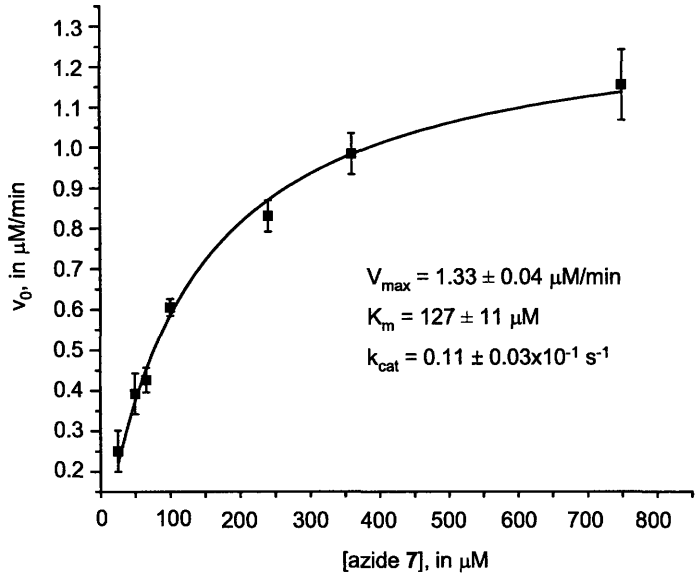


Figure 6-8: Kinetic characterization of LplA-catalyzed ligation of azide 7 to E2p. Michaelis-Menten plot for azide **7** ligation to E2p. Initial rates are shown as a function of azide **7** concentration. LplA concentration was 200 nM. From the fitted value of V_{max}, we calculated a k_{cat} value of $0.11 \pm 0.03 \times 10^{-1}$ s^{-1}. Each data point represents the average of three experiments. Error bars, 1s.d.

Rational design of a peptide substrate for LplA

For the second stage of engineering, we wished to design a peptide substrate for LplA to replace the protein substrates. It was necessary for the peptide to be fully transposable (recognized when fused to the N- or C-terminal ends of any protein) and to be recognized by LplA with similar efficiency to the natural protein substrates.

To rationally design a peptide substrate, we examined the NMR structure of the E2p domain, which presents the lysine lipoylation site at the tip of a sharp β-turn[37] (Figure 6-9). Although we acknowledged that this conformation was difficult to recapitulate in a short peptide, several studies indicated that reducing the size of E2p could indeed be possible. Initial mutagenesis studies had suggested that, while accurate

positioning of the target lysine within the β-turn was essential for LplA recognition, the residues flanking the lysine could be varied[38]. In contrast, it was later shown that deletion of a surface loop located close to the lipoyl-lysine β-turn (colored magenta in Figure 6-9) renders the domain incapable of reductive acylation, while leaving it still capable of being lipoylated, albeit at a reduced rate[39]. Importantly, NOESY analysis of the mutant domain confirmed that it was partially unfolded, with no clear lipoyl-lysine β-turn. These data suggested that the positioning of the lysine within the sharp β-turn may not be as essential for LplA recognition as it had been initially proposed. The same effect was also observed for *E. coli* E2o[40] and the E2p subunit from *Bacillus stearothermophilus*[41].

Figure 6-9: Solution structure of a hybrid lipoyl domain from *E. coli* E2p. The E2p domain folds into a flattened antiparallel β-barrel consisting of two 4-stranded β-sheets. The lipoylated lysine residue is prominently displayed in a highly conserved DKA sequence at the tip of a single β-turn in the β-sheet opposite to the N- and C-terminal residues. The lysine site of lipoylation is represented in ball and stick format. Carbon atoms are colored grey, oxygen atoms are colored red, nitrogen atoms are colored blue, and sulfur atoms are colored yellow. The residues forming the surface loop essential for reductive acylation are colored magenta. Residues incorporated in the rationally-designed peptides are colored orange.

Table 6-1 summarizes the amino acid sequences of the peptides tested (the lysine site of lipoylation is underlined). We designed three candidate peptides (**4**, **6**, and **7**) by copying the 17-amino acid stretch encompassing the lysine lipoylation sites in each of the three *E. coli* natural substrate proteins E2p, E2o, and H-protein. Three more peptides were designed to resemble the natural protein substrate (BCCP, peptides **2** and **3**) or the artificial peptide substrate (AP, peptide **1**) of the mechanistically related enzyme *E. coli*

biotin ligase (BirA). The *E. coli* biotin carboxyl carrier protein (BCCP) is the only natural protein substrate of BirA. BCCP and E2p have similar overall folds, with similar positioning of their acceptor lysines within β-turns[38]. It had been previously reported that substitution of the consensus MKM biotinylation sequence in BCCP by the lipoylation triad DKA resulted in reduced biotinylation of BCCP, accompanied by gain of certain ability to be lipoylated by LplA[42]. We thus designed a peptide based on the amino acid sequence of the mutant BCCP with the MKM motif replaced by DKA (peptide **2**). Later, Reche and Perham showed that further mutation of residues within the biotinyl-lysine turn of BCCP from the original MKMMNQ to the E2p-like DKASME sequence complete abolished recognition by BirA, while increasing its lipoylation rate by ~ 100-fold[38]. Based on this information, we designed peptide **3**. We also tested a peptide derived from E2p of *Bacillus stearothermophilus* (peptide **5**), as it had been previously shown to be recognized by *E. coli* LplA[43].

Table 6-1: Engineering a peptide substrate for LplA. A series of peptides (**1–10**) were cloned as N- or C-terminal fusions to the histone protein 1 (HP1). The extent of conversion in an azide **7** ligation reaction was assessed by HPLC assay and compared to that of the best peptide (peptide **8**). All measurements were performed in triplicate. Results are expressed as mean ± s.d.

Peptide	Sequence	Derived from	Fused to	% Conversion
1	GLNDIFEADKAEWHE	BirA AP	N term.	< 5
2	GDTLCIVEADKAMNQIE	*E. coli* BCCP	N term.	< 5
3	GDTLCIVEADKASMEIP	*E. coli* BCCP	N term.	53.9 ± 1.9
4	EQSLITVEGDKASMEVP	*E. coli* E2p	N term.	< 5
5	DDVLCEVQNDKAVVEIP	*B. stearoth.* E2p	N term.	11.9 ± 3.1
6	DEVLVEIETDKVVLEVP	*E. coli* E2o	N term.	99.0 ± 7.5
7	GDDCAVAESVKAASDIY	*E. coli* H-protein	N term.	< 5
8	DEVLVEIETDKAVLEVP	*E. coli* E2o	N term.	100 ± 8
9	DEVLVEIETDKAVLEVP	*E. coli* E2o	C term.	18.9 ± 2.4
10	DEVLVEIETDKAVLEVPGGEEE	*E. coli* E2o	C term.	51.7 ± 3.5

Peptides **1–7** were cloned as N-terminal fusions to the histone protein 1 (HP1)[44], a protein that our lab can express in high quantities and that we had previously used to determine the biotinylation kinetics of an acceptor peptide of the yeast biotin ligase[45]. We

expressed and purified the fusions, and we tested them using our HPLC assay. Table 6-1 summarizes the results. We observed large differences in the extent of lipoylation among the different peptides. We were surprised to find, for example, that the peptide derived from *E. coli* E2p (peptide **4**) was not recognized by LplA to a detectable level. In contrast, three other peptides (**3**, **5**, and **6**) were significantly lipoylated by wild-type LplA, with peptide **6** derived from *E. coli* E2o giving the best conversion. It is worth noting that peptide **5** was derived not from a lipoate-acceptor domain, but rather from the mutant biotinyl domain.

Next, we attempted to further improve peptide **6** by introducing point mutations. We compared the amino acid sequences of the three peptides that were lipoylated (**3**, **5** and **6**), and we incorporated those residues that were different in the sequence of peptide **6**. The results are illustrated in Figure 6-10. Only one of the point mutants, the Val(+1)Ala mutant (peptide **8**), gave an increase in ligation rate (Table 6-1).

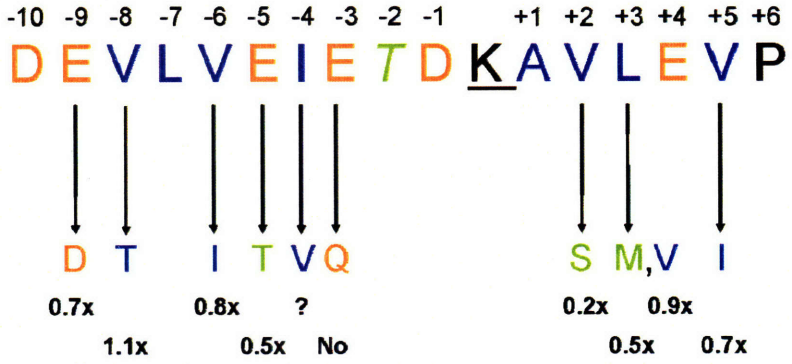


Figure 6-10: Mutagenesis analysis of an LplA peptide recognition sequence. Amino acid residues of peptide **6** were mutated to the corresponding residues present in peptides **3** or **5**. For example, Leu(+3) from peptide **6** was mutated to either Met (peptide **3**) or Val (peptide **5**). The resulting change in the lipoylation rate is expressed as a fold (x) difference with respect to the original peptide **6** sequence. Acidic amino acids are colored orange, non-polar amino acids are colored blue, and polar uncharged residues are colored green. The strictly conserved Lys and Pro(+6) were not mutated and are colored black.

To test the transposability of peptide **8**, we fused it to the C-terminal end of HP1 (to give peptide **9**). The lipoylation rate dropped 5-fold, but was then partially recovered by adding five extra C-terminal amino acids to give peptide **10**. Further addition of extra amino acids at the N- or C-terminus of peptide **10** did not affect the extent of lipoylation.

This final 22-amino acid sequence was named the LplA acceptor peptide, or LAP, and was used in all subsequent experiments.

Finally, we measured the kinetics for LplA-catalyzed ligation of azide **7** to peptide **8**. We calculated an initial rate of $0.048 \pm 0.001\ s^{-1}$ (Figure 6-6), which is only 2.3-fold slower than the corresponding rate for E2p ($0.111 \pm 0.003\ s^{-1}$, Figure 6-8). However, we estimate that the affinity of LplA for the peptide is much lower than its affinity for E2p.

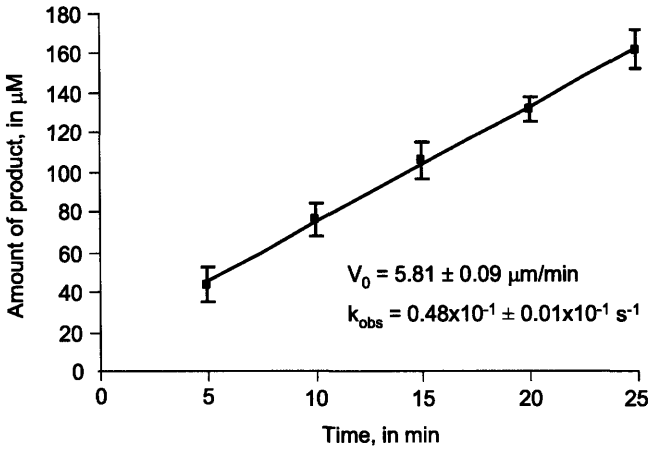


Figure 6-11: Time course of azide 7 ligation to LAP using 2 μM LplA. Reactions were incubated at 30 °C under the following reaction conditions: 1.3 mM LAP, 2 μM LplA, 750 μM azide **7**, 1 mM ATP, and 2 mM magnesium acetate in 25 mM sodium phosphate pH 7.0. Aliquots were removed at the indicated times, quenched with EDTA, and analyzed by HPLC. Each data point represents the average of three experiments. Error bars, 1 s.d.

LplA does not recognize any endogenous mammalian proteins

Finally, for the third stage of engineering, we asked whether LplA retains its high sequence specificity towards its engineered peptide substrate over all the other endogenous mammalian proteins. Previous studies had shown that protein substrates from bacteria, archaea, and plant organisms are lipoylated when recombinantly expressed in *E. coli*[43,46-48]. In contrast, protein substrates from mammalian LPLs have been shown to either not be lipoylated at all (e.g., bovine E2b)[49] or to a barely detectable extent (e.g., human E2p)[50] in *E. coli*. In the latter study, Quinn and co-workers showed that human E2p can be lipoylated *in vitro* by LplA under forcing conditions, albeit at a rate ~ 0.005-

fold lower than its *E. coli* counterpart[50]. We thus decided to perform these tests ourselves using azide **7** instead of lipoic acid.

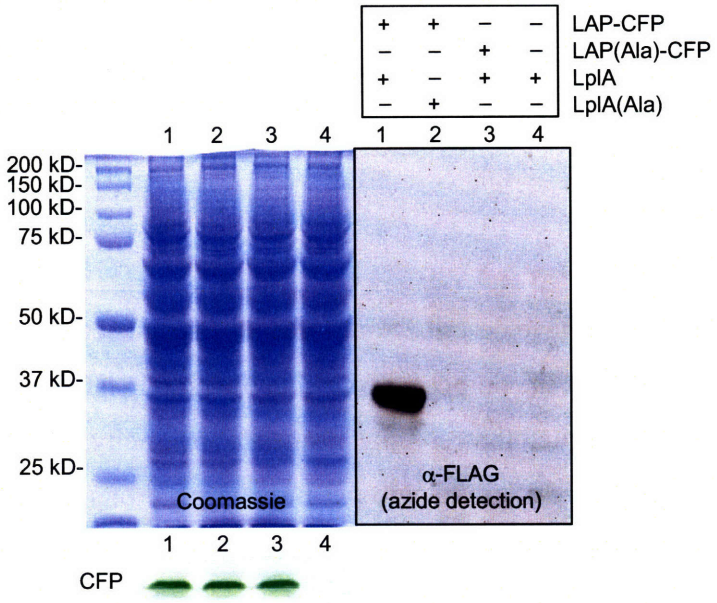


Figure 6-12: Specificity test for the ligation of azide 7 to recombinant LAP-CFP in mammalian cell lysates. Lysates from HEK cells expressing a LAP fusion to CFP were labeled *in vitro* with LplA and azide **7**. The azide was derivatized with phosphine-FLAG via the Staudinger ligation, and the FLAG epitope was detected by blotting with anti-FLAG antibody. Controls are shown with LAP-CFP replaced by its alanine point mutant (lane 3), or with LplA replaced by its catalytically inactive Lys133Ala mutant (lane 2). Coomassie staining demonstrates equal loading in all lanes. Fluorescence visualization of CFP demonstrates equal expression levels of the LAP fusion in lanes 1–3.

To test the sequence specificity of LplA, and hence the suitability for mammalian cell labeling applications, we tested LplA activity on mammalian lysates. We created a LAP fusion to cyan fluorescent protein (CFP), and expressed it in human embryonic kidney (HEK) cells. HEK lysates were then labeled with LplA, azide **7** and ATP, and the ligated azide was detected by western blot analysis, after functionalization with a FLAG peptide through the Staudinger ligation[24]. Figure 6-12 shows the anti-FLAG stained western blot. In the presence of thousands of mammalian proteins in the cell lysate, only LAP-CFP was labeled by LplA (lane 1). The expression level of LAP-CFP was so low that it could not be seen above that of endogenous proteins in the Coomassie-stained gel. Lysates from untransfected cells showed no detectable labeling (lane 4). To test the site-specificity of the reaction, we created a mutant of LAP-CFP with the lysine modification site mutated to alanine (LAP(Ala)-CFP). Lysates of cells expressing this construct,

instead of LAP-CFP, showed no labeling with FLAG (lane 3). We also performed a control in which a catalytically inactive mutant of LplA (Lys133Ala) replaced the wild-type LplA in the *in vitro* lysate labeling reaction. Lane 2 of Figure 6-12 shows that LAP-CFP labeling in lysate also depends on LplA enzymatic activity.

Conclusions

In this chapter, we have described our three-stage engineering process to re-direct *E. coli* lipoic acid ligase for the purpose of site-specific protein labeling. In previous work in our laboratory, we exploited the very high sequence specificity of *E. coli* BirA to incorporate a ketone isostere of biotin onto AP-fused target proteins[17]. Because we found that BirA is also highly specific for the biotin structure, however, many of the other biotin analogs that we synthesized and tested, including azide analogs, were rejected by the enzyme. This finding limited the usefulness of BirA for protein labeling applications. In contrast, we have shown that LplA exhibits very high specificity for its peptide or protein substrate, but relaxed specificity for its small molecule substrate. This property allowed us to incorporate alkyl azide and alkyne functional group handles, amenable for bioorthogonal ligation. To create a minimally invasive tag to direct the ligation of the alkyl azide, we engineered, through iterative cycles of rational design, a 22-amino acid replacement for LplA's natural protein substrates, which can be fused to the N- or C-terminus of any protein of interest. The successful compliance with the three aforementioned requirements (i.e., acceptance of alkyl azides, recognition of a small peptide, and high sequence specificity) makes LplA a useful platform for protein labeling applications. As explained in the introduction to this chapter, we can now use the strain-promoted [3+2] cycloaddition to site-specifically label proteins according to the two-step protein labeling procedure depicted in Figure 6-13. In the following Chapters 7 and 8, we will describe how we have used this two-step protocol to label proteins both, at the cell-surface and in the cytosol of mammalian cells.

More recently, Dr. Hemanta Baruah, a post-doctoral fellow in our lab, has found a mutant of LplA (W37I) that can efficiently ligate a small fluorophore probe based on

7-hydroxycoumarin (Figure 6-2B). In Chapters 7 and 8, we will show that direct ligation of a fluorophore by the W37I mutant results in faster labeling of cell-surface and cytosolic proteins.

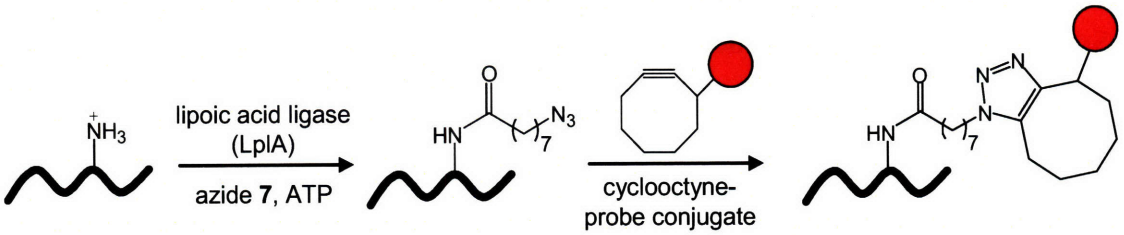


Figure 6-13: Two-step protein labeling using lipoic acid ligase. In the first step of the reaction, the *E. coli* enzyme lipoic acid ligase (LplA) catalyzes the ATP-dependent ligation of an azide-bearing analog of lipoic acid (azide **7**) to the lysine chain in the LplA acceptor peptide (LAP). In the second step of the labeling reaction, the azide undergoes [3+2] cycloaddition with a cyclooctyne-probe conjugate (red) to form a stable cycloadduct.

Experimental

General synthetic methods

Reagents were purchased from Sigma-Aldrich, Alfa Aesar, TCI America, Invitrogen, or GE Healthcare and used without further purification. Analytical thin layer chromatography was performed using 0.25 mm silica gel 60 F_{254} plates and visualized with ninhydrin or bromocresol. Flash column chromatography was carried out using silica gel (ICN SiliTech 32-63D). Mass spectra were recorded on an Applied Biosystems 200 QTRAP Mass Spectrometer using electrospray ionization. HPLC was performed on a Varian Prostar Instrument equipped with an autosampler and photo-diode-array detector. For analytical HPLC, a reverse-phase 250 × 4.6 mm Microsorb-MV 300 C18 column was used. For preparative HPLC, a reverse-phase 250 × 10 mm Microsorb-MV 100 C18 column was used. Chromatograms were recorded at 210 nm unless otherwise noted. ^{1}H NMR spectra were recorded on a Varian Mercury 300 MHz instrument. Chemical shifts are reported in delta (δ) units, parts per million (ppm), and referenced to the residual

solvent peak. Coupling constants (J) are reported in hertz (Hz). The following abbreviations for multiplets are used: s, singlet; dt, doublet of triplets; t, triplet; m, multiplet. Probes were stored as DMF solutions (3–5 M) at –20 °C.

Synthesis of alkyl azide probes

Synthesis of 7-azidoheptanoic acid[32]

To an ice-cooled solution of sodium azide (1.78 g, 27.5 mmol) in water (4.5 mL) was added dichloromethane (7.5 mL). The biphasic mixture was stirred vigorously and trifluoromethanesulfonic anhydride (0.93 mL, 5.5 mmol) was added slowly over 5 minutes. The reaction was allowed to proceed for 2 hours at 4 °C. The aqueous layer was then separated from the organic phase and extracted twice with dichloromethane (2 × 25 mL). The combined dichloromethane extracts were washed with saturated sodium carbonate solution and concentrated to 13 mL under reduced pressure. This crude triflyl azide was used without further purification. 7-aminoheptanoic acid (0.35 g, 2.4 mmol) was dissolved in water (7.8 mL) and combined with potassium carbonate (0.52 g, 3.8 mmol) and copper sulfate pentahydrate (6.6 mg, 26.4 μmol). Methanol (15.6 mL) was added to dissolve the mixture, followed by crude triflyl azide. The reaction was allowed to proceed at 21 °C overnight. The mixture was concentrated under reduced pressure, treated with NaH_2PO_4 buffer (45 mL, 250 mM, pH 6.2) and extracted with ethyl acetate (4 × 50 mL) to remove byproduct sulfonamide and excess triflyl azide. The pH of the aqueous solution was further reduced to 2.0 using concentrated HCl. The product was extracted with ethyl acetate (4 × 50 mL), dried over magnesium sulfate, and finally evaporated to dryness under reduced pressure to afford the desired product as a pale yellow oil (42 mg, 0.26 mmol, 10.7%). ^{1}H NMR ($CDCl_3$): δ 3.27 (t, 2H, J = 6.9), 2.37 (t, 2H, J = 7.5), 1.62 (m, 4H), 1.39 (m, 4H). ESI-MS calculated for $[M - H]^-$: 170.22; observed 170.16.

General procedure for synthesis of n = 5, 7–10 azides[31]

To a solution of the corresponding bromo-alkanoic acid (10 mmol of 6, 8, 9, 10 or 11-bromo-alkanoic acid) in DMF (20 mL) was added sodium azide (0.98 g, 15 mmol).

The mixture was allowed to stir at 21 °C overnight. The progress of the reaction was monitored by thin layer chromatography (0–40% ethyl acetate in hexanes). DMF was removed under reduced pressure and the resulting residue was re-suspended in HCl (25 mL, 1 N) and subsequently extracted with ethyl acetate (4 × 25 mL). The combined organic layers were dried over magnesium sulfate, and evaporated to dryness under reduced pressure to afford the desired azido-alkanoic acid as a pale yellow oil. Typical yields ranged from 35–65%.

Characterization data for n = 5 azide (6-azidohexanoic acid). ^{1}H NMR ($CDCl_3$): δ 3.28 (t, 2H, J = 6.9), 2.37 (t, 2H, J = 7.5), 1.64 (m, 4H), 1.35 (m, 2H). ESI-MS calculated for $[M - H]^-$: 156.17; observed 156.12.

Characterization data for n = 7 azide (8-azidooctanoic acid). ^{1}H NMR ($CDCl_3$): δ 3.26 (t, 2H, J = 6.8), 2.36 (t, 2H, J = 7.5), 1.62 (m, 4H), 1.35 (m, 6H). ESI-MS calculated for $[M - H]^-$: 184.22; observed 184.14.

Characterization data for n = 8 azide (9-azidononanoic acid). ^{1}H NMR ($CDCl_3$) δ 3.25 (t, 2H, J = 6.9), 2.34, (t, 2H, J = 7.5), 1.60 (m, 4H), 1.32 (m, 8H). ESI-MS calculated for $[M - H]^-$: 198.25; observed 198.18.

Characterization data for n = 9 azide (10-azidodecanoic acid). ^{1}H NMR ($CDCl_3$) δ 3.25 (t, 2H, J = 7.1), 2.35, (t, 2H, J = 7.5), 1.61 (m, 4H), 1.31 (m, 10H). ESI-MS calculated for $[M - H]^-$: 212.28; observed 212.16.

Characterization data for n = 10 azide (11-azidoundecanoic acid). ^{1}H NMR ($CDCl_3$) δ 3.25 (t, 2H, J = 7.1), 2.35, (t, 2H, J = 7.5), 1.62 (m, 4H), 1.29 (m, 12H). ESI-MS calculated for $[M - H]^-$: 226.31; observed 226.20.

Synthesis of alkyne probes

Synthesis of n = 5 and n = 7 alkynes[33]

To a solution of 10 mmol 6- or 8-bromo-alkanoic acid in DMSO (23 mL) was added lithium acetylide ethylenediamine (0.92 g, 10 mmol) slowly over a period of 5 minutes. The reaction mixture was stirred overnight at 21 °C. Subsequently, water (25 mL) was added and the product extracted with dichloromethane (3 × 25 mL).

Dichloromethane was removed under reduced pressure to yield the desired alkynoic acid as a brownish solid. Yields ranged from 5–7%.

*Characterization data for n = 5 alkyne (*7-octynoic acid*).* ^{1}H NMR ($CDCl_3$) δ 2.36 (t, 2H, J = 7.1), 2.19, (t, 2H, J = 6.9), 1.94 (t, 1H, J = 2.6), 1.69–1.34(m, 6H). ESI-MS calculated for $[M - H]^-$: 139.18; observed 139.08.

*Characterization data for n = 7 alkyne (*9-decynoic acid*).* ^{1}H NMR ($CDCl_3$) δ 2.35 (t, 2H, J = 7.5), 2.18, (dt, 2H, J = 6.9, J = 2.7), 1.94 (t, 1H, J = 2.7), 1.63 (m, 2H), 1.51 (m, 2H), 1.35 (m, 6H) ESI-MS calculated for $[M - H]^-$: 167.23; observed 167.22.

Cloning of HP1-peptide fusions for recombinant expression in bacteria

For peptides **1-7**, the histone protein 1 (HP1) gene[44] was PCR-amplified using a forward primer that introduced the desired peptide sequence after an NheI site (forward primer sequences below), and the reverse primer HP1-EcoRI.R (5′ TTTT *GAA TTC* GGA TCC TTG CGG CTC GCC TCG TAC). The resulting PCR product was digested with *NheI* and *EcoRI* and ligated in-frame to an *NheI/EcoRI* digested pET21b vector. The vector introduced a C-terminal His_6 tag.

Peptide	Forward Primer Sequence (NheI sites are italicized)
1	5′ AAAA *GCT AGC* GGC CTG AAC GAC ATC TTC GAA GCC GAC AAA GCT GAA TGG CAC GAG GGC GGT GAG GAG GAG TAC GCC GTG G
2	5′ AAAA *GCT AGC* GGC GAT ACC CTG TGC ATC GTT GAA GCC GAC AAA GCT GAA AAC CAG ATC GAA GGC GGT GAG GAG GAG TAC GCC GTG G
3	5′ AAAA *GCT AGC* GGC GAT ACC CTG TGC ATC GTT GAA GCC GAC AAA GCT TCT ATG GAA ATC CCG GGC GGT GAG GAG GAG TAC GCC GTG G
4	5′ AAAA *GCT AGC* GAA CAG TCG CTG ATC ACC GTA GAA GGC GAC AAA GCT TCT ATG GAA GTT CCG GGC GGT GAG GAG GAG TAC GCC GTG G
5	5′ AAAA *GCT AGC* GAC GAT GTA CTG TGC GAA GTA CAG AAC GAC AAA GCT GTA GTT GAA ATC CCG GGC GGT GAG GAG GAG TAC GCC GTG G
6	5′ AAAA *GCT AGC* GAC GAA GTA CTG GTT GAA ATC GAA ACC GAC AAA GTA GTT CTG GAA GTA CCG GGC GGT GAG GAG GAG TAC GCC GTG G
7	5′ AAAA GCT AGC GGC GAT GAC TGC GCT GTT GCT GAA TCT GTA AAA GCT GCC TCG GAC ATC TAT GGC GGT GAG GAG GAG TAC GCC GTG G

Peptide **8** was obtained by mutagenesis on peptide **6**, using the QuikChange primer 5′ GAA ATC GAA ACC GAC AAA GCA GTT CTG GAA GTA CCG GGC and its reverse complement. To clone peptide **9**, the HP1 gene was PCR-amplified using the primer NdeI-HP1.F (5′ AAAA *CAT ATG* GAG GAG GAG TAC GCC GTG G), which incorporates an *NdeI* site, and the primer HP1-LAP-Stop-BamHI.R (5′ TTTT *GGA TCC* TCT TAC GGT ACT TCC AGA ACT GCT TTG TCG GTT TCG ATT TCA ACC AGT ACT TCG TCG CTA GCA TCC TTG CGG CTC GCC TCG TAC), which introduces the peptide sequence and a *BamHI* site. The resulting PCR product was digested with *NdeI* and *BamHI* and ligated in-frame to an *NdeI/BamHI* digested pET15b vector, which introduces an N-terminal His_6 tag. To clone peptide **10**, the gene encoding peptide **9** was PCR-amplified using the same forward primer NdeI-HP1.F, and the reverse primer LAP-AAs-Stop-BamHI.R (5′ TTTT *GGA TCC* TCT TAC TCC TCC TCA CCG CCC GGT ACT TCC AGA ACT GCT TTG TC). The PCR product was digested and ligated as above.

Cloning of LAP-CFP for cytoplasmic mammalian expression

The sequence for the LAP peptide was inserted in the *NheI* site of a modified form of the pcDNA3 vector (Invitrogen). The vector contained the CFP gene between the *BamHI* and *EcoRI* sites[17]. Primers 5′ AAAA *ACT AGT* CGG GCT GAC GAA GTA CTG GTT GAA ATC GAA ACC GAC AAA GCA GTT CTG GAA GTA CCG GCA TCA GCA GAC GGC *GCT AGC* AAAA and its reverse complement were annealed together, digested with *SpeI* and *NheI*, and ligated in-frame to *NheI* digested pcDNA3-CFP vector. To create the LAP(Ala)-CFP mutant, we performed QuikChange with the primer 5′ G GTT GAA ATC GAA ACC GAC GCC GCA GTT CTG GAA GTA CCG G and its reverse complement.

Site directed mutagenesis of LplA

The pYFJ16 plasmid, a gift from John Cronan, encodes an N-terminally His_6-tagged LplA within the pQE-2 vector (Qiagen). To create the catalytically inactive mutant, Lys133 was mutated to alanine by QuikChange using 5′ CGA AGG CGA CCG CGC AGT CTC AGG CTC GG and its reverse complement.

Bacterial expression and purification of *E. coli* LplA

LplA was expressed from the plasmid pYFJ16, a gift from John Cronan, which encodes an N-terminally His_6-tagged LplA within the pQE-2 vector (Qiagen). pYFJ16 was transformed into *E. coli* BL21(DE3) cells, which were amplified in LB media supplemented with 100 μg/mL ampicillin at 37 °C until OD_{600} 0.9. Enzyme expression was induced with 200 μg/mL IPTG for 3 hours at 30 °C. Thereafter, cells were harvested by centrifugation (6,000 rpm, 10 minutes, 4 °C) and the pellet was resuspended in lysis buffer (50 mM Tris base, 300 mM NaCl, pH 7.8) containing 2.5 mM phenylmethylsulfonyl fluoride (PMSF) and protease inhibitor cocktail (Calbiochem). Cells were lysed by ultrasonic treatment (six 15-second bursts, with 1 minute of cooling to 4 °C between bursts). The extract was cleared by centrifugation (17,700 g, 10 minutes, 4 °C) and the His_6-tagged enzyme was purified using Ni-NTA agarose (Qiagen). Fractions were analyzed by 12% SDS-PAGE followed by Coomassie staining. Fractions containing LplA were pooled and dialyzed against PBS pH 7.4. Enzyme concentration was determined by measuring A_{280} and using the reported extinction coefficient ($46{,}250\ M^{-1}\ cm^{-1}$)[1].

Bacterial expression and purification of E2p

A single hybrid lipoyl domain derived from the second subunit of the *E. coli* pyruvate dehydrogenase was expressed from the *E. coli* K12 strain, TM245, which was a gift from John Cronan[7,34]. Transformants of strain TM245 were grown at 37 °C in LB media supplemented with 100 μg/mL ampicillin until OD_{600} 0.2. Protein expression was induced with 10 μg/mL IPTG for 17 hours at 25 °C. Harvested bacteria were resuspended in 20 mM sodium phosphate buffer pH 7.4 containing 2 mM EDTA, 2.5 mM PMSF, and protease inhibitor cocktail (Calbiochem). Cells were lysed by ultrasonic treatment (eight 30-second bursts, with 30 seconds of cooling to 4 °C between bursts). The extract was cleared by centrifugation (17,700 g, 40 minutes, 4 °C), before lowering the pH to 3.8 with 1 M HCl. Insoluble material was removed by two rounds of centrifugation (17,700 g, 20 minutes, 4 °C and 17,700 g, 10 minutes, 4 °C), before increasing the pH to 7.0 with 1 M NaOH. The supernatant was dialyzed against 10 mM ammonium acetate pH 5.0,

then subjected to fast flow anion-exchange chromatography on a 1 mL Q-Sepharose column with a 10–600 mM ammonium acetate pH 5.0 gradient generated over 20 column volumes. Eluted fractions were analyzed by 19% SDS-PAGE followed by Coomassie staining. Fractions containing E2p were pooled and dialyzed against PBS pH 7.4. Protein concentration was measured using the BCA (bicinchoninic acid) assay with BSA as the standard.

Bacterial expression and purification of HP1-peptide fusions

E. coli BL21(DE3) cells transformed with one of the HP1 expression plasmids were amplified in LB media supplemented with 100 μg/mL ampicillin at 37 °C until OD_{600} 0.9. Protein expression was induced with 100 μg/mL IPTG for 4 hours at 30 °C. Cells were harvested by centrifugation and purified as described above for LplA. Purification fractions were analyzed by 16% SDS-PAGE followed by Coomassie staining. Fractions containing HP1-peptide protein were pooled and dialyzed against PBS pH 7.4. Protein concentration was measured using the BCA assay with BSA as the reference standard.

***In vitro* LplA activity assays**

LplA reactions contained 2 μM LplA, 200 μM E2p, 350 μM probe, 1 mM ATP, 2 mM magnesium acetate, and 25 mM sodium phosphate pH 7.0. Reactions were incubated at 30 °C for 30 minutes, and then quenched with EDTA (final concentration 50 mM). Conversion to product was determined by HPLC on a C_{18} reverse-phase column with a 40–57% gradient of acetonitrile in water with 0.1% trifluoroacetic acid over 20 minutes (flow rate 1.0 mL/minute). Unmodified E2p had a retention time of ~ 12 minutes while E2p-probe conjugates eluted at 15–18 minutes. Percent conversion to product was calculated from the ratio of the E2p-probe peak area to the sum of (E2p + E2p-probe) peak areas. All measurements were performed in triplicate.

Measurement of LplA-probe ligation kinetics

To measure the k_{cat} for LplA ligation of lipoic acid to E2p, 50 nM LplA was combined with 200 μM E2p, 750 μM lipoic acid, 1 mM ATP, and 2 mM magnesium acetate in 25 mM sodium phosphate pH 7.0. The reaction was initiated with addition of 1 mM ATP to the pre-warmed (30 °C) mixture. The reaction was incubated at 30 °C, and 90 μL aliquots were removed every 5 minutes and quenched with 50 mM EDTA (final concentration). Samples were analyzed by C18 reverse-phase HPLC as described in the "Methods" section. Measurements were performed in triplicate. A calibration curve was obtained that correlated the ratio of integrated HPLC peak areas (E2p : E2p-probe) to the actual protein ratio, in order to compensate for differences in the extinction coefficients of E2p and lipoylated E2p. The amount of product obtained at each time point was plotted against time, to obtain the V_{max} of the reaction. From this, the k_{cat} value was obtained using the equation $V_{max} = k_{cat}\,[E]_{total}$.

To measure the K_m and k_{cat} for ligation of azide **7**, 200 nM LplA was combined with 200 μM E2p, 1 mM ATP, 2 mM magnesium acetate in 25 mM sodium phosphate pH 7.0, and various concentrations of azide **7** (25, 50, 66, 100, 240, 360 and 750 μM). 90 μL aliquots were removed from the 30 °C reactions at 5 minutes intervals, up to 30 minutes, and quenched with 50 mM EDTA (final concentration). Samples were analyzed by HPLC as above. The amount of product obtained at each time point was plotted against time, to obtain the initial velocity for each concentration of azide **7**. The collection of initial velocities (V_0) was then plotted against azide **7** concentration (25–750 μM), and fit to the Michaelis-Menten equation ($V_0 = V_{max}$ [azide **7**]/(K_m + [azide **7**])), using Origin 6.1 software, to obtain the K_m for azide **7**. From the V_{max} value, k_{cat} was calculated using the equation $V_{max} = k_{cat}\,[E]_{total}$.

Mass-spectrometric analysis of E2p-azide 7 conjugate

2 μM LplA was combined with 200 μM E2p, 350 μM azide **7**, 1 mM ATP, and 2 mM magnesium acetate in 25 mM sodium phosphate pH 7.0. The ligation reaction was allowed to proceed to completion by incubating at 30 °C for 2 hours. The reaction mixture was desalted by extensive dialysis against water (2 × 4 h, followed by overnight

dialysis). Thereafter, the sample was diluted to a final concentration of 25 μM E2p-azide **7** in 50% methanol with 2% acetic acid. Mass spectra were recorded under the positive enhanced multi-charge mode of an ESI-MS.

HPLC analysis of LplA modification of HP1-peptide fusions

To compare the conversion rates for different peptide substrates, reactions were assembled as follows: 1.5 μM LplA, 150 μM HP1-peptide fusion, 750 μM azide **7**, 1 mM ATP, and 2 mM magnesium acetate in 25 mM sodium phosphate pH 7.0. Reactions were incubated at 30 °C for 2 hours, then quenched with 50 mM EDTA (final concentration), and subsequently analyzed by C18 reverse-phase HPLC using a gradient of 30–45% acetonitrile in water with 0.1% trifluoroacetic acid over 20 minutes with a 1 mL/minute flow rate. Retention times for unmodified HP1-peptide fusions ranged from 8–12 minutes, and shifted to 16–21 minutes after ligation to azide **7**. Measurements were performed in triplicate. The extent of modification was calculated from the ratio of HP1-peptide-azide **7** peak area to the sum of (HP1-peptide + HP1-peptide-azide **7**) peak areas. The conversion with HP1-peptide 8 was normalized to 100% and other conversions were reported relative to this.

Measurement of k_{cat} for LplA-catalyzed azide 7 ligation to LAP

To measure the k_{cat} for LplA-catalyzed ligation of azide **7** to LAP, the reaction conditions were as follows: 2 μM LplA, 1.3 mM LAP-HP1 (N-terminal fusion of peptide 10 to HP1), 750 μM azide **7**, 1 mM ATP, and 2 mM magnesium acetate in 25 mM sodium phosphate pH 7.0. The reaction was initiated with addition of 1 mM ATP to the pre-warmed 30 °C mixture. The mixture was incubated at 30 °C, and 10 μL aliquots were removed every 5 minutes and quenched with 50 mM EDTA (final concentration). Samples were analyzed by C18 reverse-phase HPLC as above. The amount of product obtained at each time point was plotted against time, to obtain the V_{max} of the reaction. From this, the k_{cat} value was obtained using the equation $V_{max} = k_{cat}\,[E]_{total}$.

LplA specificity test on mammalian lysates

Human embryonic kidney (HEK) 293T cells were transfected with LAP-CFP-pcDNA3 plasmid using Lipofectamine 2000 (1 μg DNA/well of a 6-well plate). Lysates were generated 48 hours later by hypotonic lysis to minimize protease release, as follows. Cells were lifted from the plates, concentrated by centrifugation, and resuspended in 1 mM HEPES pH 7.5, 5 mM magnesium chloride, 1 mM phenylmethylsulphonyl fluoride, and protease inhibitor cocktail (Calbiochem). After incubation at 4 °C for 10 minutes, the cells were lysed by vigorous vortexing for 2 minutes at 21 °C. Crude lysate was clarified by centrifugation, and stored at –80 °C. Lysate was labeled by incubating at 30 °C for 10 hours with 25 mM sodium phosphate pH 7.0, 1 μM LplA, 250 μM azide **7**, 1 mM ATP, and 4 mM magnesium acetate. Thereafter, Staudinger ligation was performed by adding FLAG-phosphine[24] to a final concentration of 500 μM, and incubating at 30 °C for 16 hours. Each reaction sample was then divided into thirds. The first third was analyzed by 12% SDS-PAGE followed by Western blotting with anti-FLAG(M2)-peroxidase antibody conjugate (Sigma, 1:1000 dilution). The second sample was analyzed by 12% SDS-PAGE followed by Coomassie staining. The last third was analyzed by 12% SDS-PAGE without boiling the samples, in order to prevent unfolding of CFP, and in-gel fluorescence was visualized on a Storm 860 instrument (Amersham).

References

1. Green,D.E., Morris,T.W., Green,J., Cronan,J.E., Jr. & Guest,J.R. Purification and properties of the lipoate protein ligase of Escherichia coli. *Biochem. J.* **309 (Pt 3)**, 853-862 (1995).

2. Cronan,J.E., Zhao,X. & Jiang,Y. Function, attachment and synthesis of lipoic acid in Escherichia coli. *Adv. Microb. Physiol* **50**, 103-146 (2005).

3. Fujiwara,K., Okamura-Ikeda,K. & Motokawa,Y. Cloning and expression of a cDNA encoding bovine lipoyltransferase. *J. Biol. Chem.* **272**, 31974-31978 (1997).

4. Thekkumkara,T.J., Pons,G., Mitroo,S., Jentoft,J.E. & Patel,M.S. Molecular biology of the human pyruvate dehydrogenase complex: structural aspects of the E2 and E3 components. *Ann. N. Y. Acad. Sci.* **573**, 113-129 (1989).

5. Fujiwara,K. *et al.* Molecular cloning, structural characterization and chromosomal localization of human lipoyltransferase gene. *Eur. J. Biochem.* **260**, 761-767 (1999).

6. Fujiwara,K., Takeuchi,S., Okamura-Ikeda,K. & Motokawa,Y. Purification, characterization, and cDNA cloning of lipoate-activating enzyme from bovine liver. *J. Biol. Chem.* **276**, 28819-28823 (2001).

7. Morris,T.W., Reed,K.E. & Cronan,J.E., Jr. Lipoic acid metabolism in Escherichia coli: the lplA and lipB genes define redundant pathways for ligation of lipoyl groups to apoprotein. *J. Bacteriol.* **177**, 1-10 (1995).

8. Jordan,S.W. & Cronan,J.E., Jr. The Escherichia coli lipB gene encodes lipoyl (octanoyl)-acyl carrier protein:protein transferase. *J. Bacteriol.* **185**, 1582-1589 (2003).

9. Miller,J.R. *et al.* Escherichia coli LipA is a lipoyl synthase: in vitro biosynthesis of lipoylated pyruvate dehydrogenase complex from octanoyl-acyl carrier protein. *Biochemistry* **39**, 15166-15178 (2000).

10. Yi,X. & Maeda,N. Endogenous production of lipoic acid is essential for mouse development. *Mol. Cell Biol.* **25**, 8387-8392 (2005).

11. Fujiwara,K. *et al.* Crystal structure of lipoate-protein ligase A from Escherichia coli. Determination of the lipoic acid-binding site. *J. Biol. Chem.* **280**, 33645-33651 (2005).

12. Fujiwara,K. *et al.* Crystal structure of bovine lipoyltransferase in complex with lipoyl-AMP. *J. Mol. Biol.* **371**, 222-234 (2007).

13. Kim,D.J. *et al.* Crystal structure of lipoate-protein ligase A bound with the activated intermediate: insights into interaction with lipoyl domains. *J. Biol. Chem.* **280**, 38081-38089 (2005).

14. McManus,E., Luisi,B.F. & Perham,R.N. Structure of a putative lipoate protein ligase from Thermoplasma acidophilum and the mechanism of target selection for post-translational modification. *J. Mol. Biol.* **356**, 625-637 (2006).

15. Morris,T.W., Reed,K.E. & Cronan,J.E., Jr. Identification of the gene encoding lipoate-protein ligase A of Escherichia coli. Molecular cloning and characterization of the lplA gene and gene product. *J. Biol. Chem.* **269**, 16091-16100 (1994).

16. Prescher,J.A. & Bertozzi,C.R. Chemistry in living systems. *Nat. Chem. Biol.* **1**, 13-21 (2005).

17. Chen,I., Howarth,M., Lin,W. & Ting,A.Y. Site-specific labeling of cell surface proteins with biophysical probes using biotin ligase. *Nat. Methods* **2**, 99-104 (2005).

18. Huisgen,R. 1,3-Dipolar Cycloadditions. Past and Future. *Angew. Chem. , Int. Ed. Engl.* **2**, 565-598 (1963).

19. Wang,Q. *et al.* Bioconjugation by copper(I)-catalyzed azide-alkyne [3 + 2] cycloaddition. *J. Am. Chem. Soc.* **125**, 3192-3193 (2003).

20. Hsu,T.L. *et al.* Alkynyl sugar analogs for the labeling and visualization of glycoconjugates in cells. *Proc. Natl. Acad. Sci. U. S. A* **104**, 2614-2619 (2007).

21. Link,A.J., Vink,M.K. & Tirrell,D.A. Presentation and detection of azide functionality in bacterial cell surface proteins. *J. Am. Chem. Soc.* **126**, 10598-10602 (2004).

22. Speers,A.E., Adam,G.C. & Cravatt,B.F. Activity-based protein profiling in vivo using a copper(i)-catalyzed azide-alkyne [3 + 2] cycloaddition. *J. Am. Chem. Soc.* **125**, 4686-4687 (2003).

23. Saxon,E. & Bertozzi,C.R. Cell surface engineering by a modified Staudinger reaction. *Science* **287**, 2007-2010 (2000).

24. Kiick,K.L., Saxon,E., Tirrell,D.A. & Bertozzi,C.R. Incorporation of azides into recombinant proteins for chemoselective modification by the Staudinger ligation. *Proc. Natl. Acad. Sci. U. S. A* **99**, 19-24 (2002).

25. Ohno,S. *et al.* Site-selective post-translational modification of proteins using an unnatural amino acid, 3-azidotyrosine. *J. Biochem.* **141**, 335-343 (2007).

26. Vocadlo,D.J., Hang,H.C., Kim,E.J., Hanover,J.A. & Bertozzi,C.R. A chemical approach for identifying O-GlcNAc-modified proteins in cells. *Proc. Natl. Acad. Sci. U. S. A* **100**, 9116-9121 (2003).

27. Agard,N.J., Baskin,J.M., Prescher,J.A., Lo,A. & Bertozzi,C.R. A comparative study of bioorthogonal reactions with azides. *ACS Chem. Biol.* **1**, 644-648 (2006).

28. Agard,N.J., Prescher,J.A. & Bertozzi,C.R. A strain-promoted [3 + 2] azide-alkyne cycloaddition for covalent modification of biomolecules in living systems. *J. Am. Chem. Soc.* **126**, 15046-15047 (2004).

29. Baskin,J.M. *et al.* Copper-free click chemistry for dynamic in vivo imaging. *Proc. Natl. Acad. Sci. U. S. A* **104**, 16793-16797 (2007).

30. Laughlin,S.T., Baskin,J.M., Amacher,S.L. & Bertozzi,C.R. In vivo imaging of membrane-associated glycans in developing zebrafish. *Science* **320**, 664-667 (2008).

31. Scriven,E.F.V. & Turnbull,K. Azides: their preparation and synthetic uses. *Chemical Reviews* **88**, 297-368 (1998).

32. Alper,P.B., Hung,S.C. & Wong,C.H. Metal catalyzed diazo transfer for the synthesis of azides from amines. *Tetrahedron Letters* **37**, 6029-6032 (1996).

33. Novis-Smith,W. & Beumel,J. Preparation of alkynes and dialkynes by reaction of mono-halo and dihaloalkanes with lithium acetylenide-ethylenediamine complex. *Synthetic Communications* 441-442 (1974).

34. Ali,S.T. & Guest,J.R. Isolation and characterization of lipoylated and unlipoylated domains of the E2p subunit of the pyruvate dehydrogenase complex of Escherichia coli. *Biochem. J.* **271**, 139-145 (1990).

35. Reed,L.J. & Hackert,M.L. Structure-function relationships in dihydrolipoamide acyltransferases. *J. Biol. Chem.* **265**, 8971-8974 (1990).

36. Miles,J.S. & Guest,J.R. Subgenes expressing single lipoyl domains of the pyruvate dehydrogenase complex of Escherichia coli. *Biochem. J.* **245**, 869-874 (1987).

37. Green,J.D., Laue,E.D., Perham,R.N., Ali,S.T. & Guest,J.R. Three-dimensional structure of a lipoyl domain from the dihydrolipoyl acetyltransferase component of the pyruvate dehydrogenase multienzyme complex of Escherichia coli. *J. Mol. Biol.* **248**, 328-343 (1995).

38. Reche,P. & Perham,R.N. Structure and selectivity in post-translational modification: attaching the biotinyl-lysine and lipoyl-lysine swinging arms in multifunctional enzymes. *EMBO J.* **18**, 2673-2682 (1999).

39. Jones,D.D., Horne,H.J., Reche,P.A. & Perham,R.N. Structural determinants of post-translational modification and catalytic specificity for the lipoyl domains of the pyruvate dehydrogenase multienzyme complex of Escherichia coli. *J. Mol. Biol.* **295**, 289-306 (2000).

40. Jones,D.D. & Perham,R.N. The role of loop and beta-turn residues as structural and functional determinants for the lipoyl domain from the Escherichia coli 2-oxoglutarate dehydrogenase complex. *Biochem. J.* **409**, 357-366 (2008).

41. Wallis,N.G., Allen,M.D., Broadhurst,R.W., Lessard,I.A. & Perham,R.N. Recognition of a surface loop of the lipoyl domain underlies substrate channelling in the pyruvate dehydrogenase multienzyme complex. *J. Mol. Biol.* **263**, 463-474 (1996).

42. Reche,P., Li,Y.L., Fuller,C., Eichhorn,K. & Perham,R.N. Selectivity of post-translational modification in biotinylated proteins: the carboxy carrier protein of the acetyl-CoA carboxylase of Escherichia coli. *Biochem. J.* **329 (Pt 3)**, 589-596 (1998).

43. Dardel,F., Packman,L.C. & Perham,R.N. Expression in Escherichia coli of a sub-gene encoding the lipoyl domain of the pyruvate dehydrogenase complex of Bacillus stearothermophilus. *FEBS Lett.* **264**, 206-210 (1990).

44. Saunders,W.S. *et al.* Molecular cloning of a human homologue of Drosophila heterochromatin protein HP1 using anti-centromere autoantibodies with anti-chromo specificity. *J. Cell Sci.* **104 (Pt 2)**, 573-582 (1993).

45. Chen,I., Choi,Y.A. & Ting,A.Y. Phage display evolution of a peptide substrate for yeast biotin ligase and application to two-color quantum dot labeling of cell surface proteins. *J. Am. Chem. Soc.* **129**, 6619-6625 (2007).

46. Berg,A., Westphal,A.H., Bosma,H.J. & de Kok,A. Kinetics and specificity of reductive acylation of wild-type and mutated lipoyl domains of 2-oxo-acid dehydrogenase complexes from Azotobacter vinelandii. *Eur. J. Biochem.* **252**, 45-50 (1998).

47. MacHerel,D. *et al.* Expression, lipoylation and structure determination of recombinant pea H-protein in Escherichia coli. *Eur. J. Biochem.* **236**, 27-33 (1996).

48. Wallis,N.G. & Perham,R.N. Structural dependence of post-translational modification and reductive acetylation of the lipoyl domain of the pyruvate dehydrogenase multienzyme complex. *J. Mol. Biol.* **236**, 209-216 (1994).

49. Griffin,T.A., Wynn,R.M. & Chuang,D.T. Expression and assembly of mature apotransacylase (E2b) of bovine branched-chain alpha-keto acid dehydrogenase complex in Escherichia coli. Demonstration of transacylase activity and modification by lipoylation. *J. Biol. Chem.* **265**, 12104-12110 (1990).

50. Quinn,J. *et al.* Expression and lipoylation in Escherichia coli of the inner lipoyl domain of the E2 component of the human pyruvate dehydrogenase complex. *Biochem. J.* **289 (Pt 1)**, 81-85 (1993).

Chapter 7 : Site-specific labeling of cell-surface proteins with lipoic acid ligase

The work discussed in this chapter has been published in part in: M. Fernández-Suárez, H. Baruah, L. Martínez-Hernández, K. T. Xie, J. M. Baskin, C. R. Bertozzi, and A. Y. Ting. Redirecting lipoic acid ligase for cell surface protein labeling with small-molecule probes. *Nat. Biotechnol.* **25**(12), 1483-1487 (2007). The cyclooctyne conjugates and coumarin probes were synthesized by Dr. Hemanta Baruah. Jeremy Baskin provided the cyclooctyne starting material.

Introduction

In Chapter 6, we demonstrated that LplA was a promising platform to develop a new protein labeling methodology. We engineered small molecule and peptide substrates for LplA and we demonstrated that LplA is highly specific for labeling its peptide substrate in the context of cell lysates. We were then ready to test the methodology for labeling proteins in the live cell context. We started by testing LplA labeling at the cell surface because it was less demanding than intracellular labeling. For example, whereas for cell-surface protein labeling, LplA can be added in solution, intracellular labeling requires that the enzyme be transfected and expressed inside cells. Additionally, cell-surface protein labeling eliminated the need to find probes capable of permeating across the cell membrane. In this chapter, we first demonstrate that LplA can be used to label LAP-tagged proteins with a variety of probes at the surface of mammalian cells, and we then apply the methodology to one- and two-color cell-surface receptor labeling. We also characterize the speed, sensitivity, and specificity of the method, and we compare it to our previously reported biotin ligase and transglutaminase methods.

Results

Site-specific labeling of LAP fusion proteins with small organic fluorophores

Building onto our finding that LplA can accept alkyl azide analogs of lipoic acid, we devised a two-step labeling methodology to biotinylate or fluorescently tag the LplA acceptor peptide (LAP) on the surface of mammalian cells. In our scheme for cell-surface labeling (Figure 7-1), the LAP peptide is expressed as a fusion to the extracellular domain of a membrane protein, and then labeled with LplA, azide **7** and ATP. In the second step of the reaction, we use the [3+2] cycloaddition reaction to introduce a variety of probes (e.g., biotin or a small organic fluorophore).

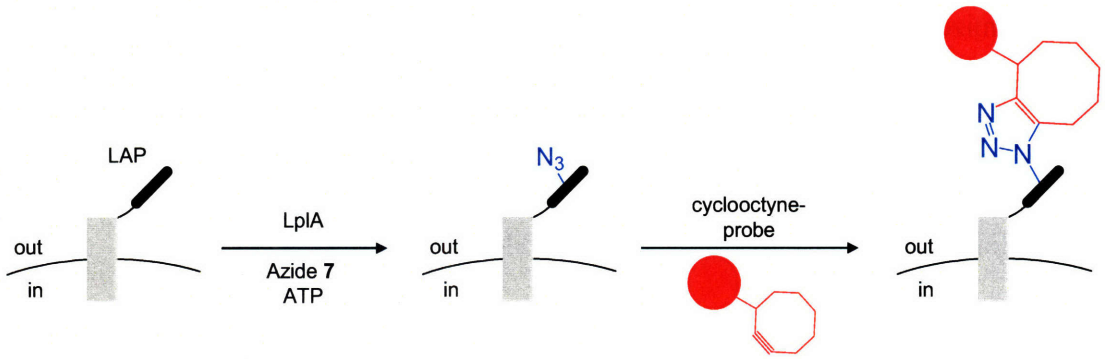


Figure 7-1: Two-step cell-surface labeling of LAP-tagged proteins. The labeling procedure is performed in two steps. In the first step, azide **7** is enzymatically ligated to cell-surface LAP. In the second step, after washing the LplA, azide **7,** and ATP reagents, a cyclooctyne conjugate chemoselectively reacts with the azide-derivatized LAP-tagged proteins.

One of the most important determinants when performing chemical labeling in live cells is the choice of appropriate probes. In the case of cell-surface protein labeling, the molecules need to be highly hydrophilic to render the molecule membrane impermeant and to avoid non-specific binding with the hydrophobic cell membrane. Hydrophilicity thus minimizes background staining from endocytosed and non-specifically bound fluorophore. We chose Alexa Fluor and cyanine dyes because they are the brightest red-emitting, impermeant fluorophores, and because they are commercially available in the form of activated esters. We used these activated esters to synthesize fluorophore conjugates of monofluoro-cyclooctyne (OCT)[1]. As previously mentioned in the introduction of Chapter 6, cyclooctynes react with azides to form stable triazoles under physiological conditions without the need for a metal catalyst[2]. Introduction of an electron-withdrawing fluorine atom alpha to the alkyne increases the second order rate of the cycloaddition reaction by about 2-fold[1]. Although it was later shown by the Bertozzi group that introduction of a second fluorine further accelerates the reaction by ~ 18-fold, those probes were not yet available at the time when we were performing these experiments. We obtained the cyclooctyne starting material (OCT acid, **1**) from Jeremy Baskin (Bertozzi lab). The synthesis of the OCT-fluorophore conjugates was performed by Dr. Hemanta Baruah, a post-doctoral fellow in our lab following the scheme in Figure 7-2. OCT acid (**1**) was first activated as a pentafluorophenyl (PFP) ester using PFP-TFA, and the OCT-PFP (**2**) was then conjugated to *O*,*O*'-Bis(3-aminopropyl)diethylene glycol (diamino-PEG) linker to give OCT-PEG (**3**). The use of PEG improves water solubility

and minimizes non-specific interactions with the cell membrane. The OCT-PEG **3** conjugate was then coupled to the *N*-hydrosuccinimidyl esters of Alexa Fluor 568 and Cy3 to give OCT-PEG-Alexa Fluor 568 (**4**) and OCT-PEG-Cy3 (**5**). We confirmed the structures of all these probes by mass spectrometry and by reaction with unconjugated azide **7** followed by detection of the triazole adduct (data not shown).

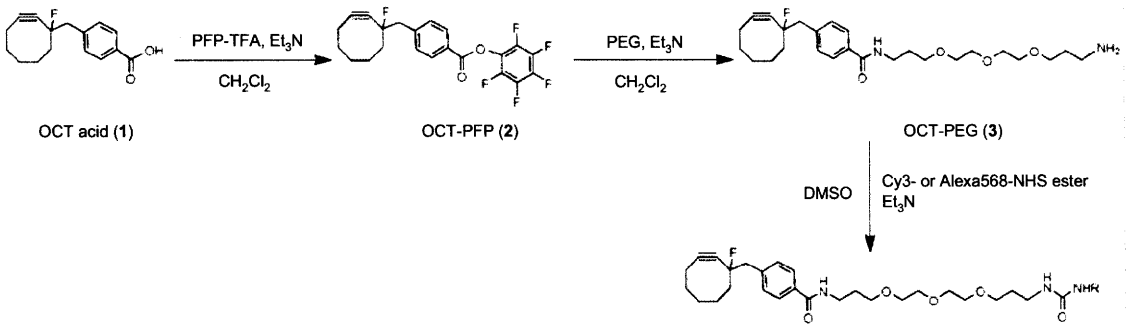
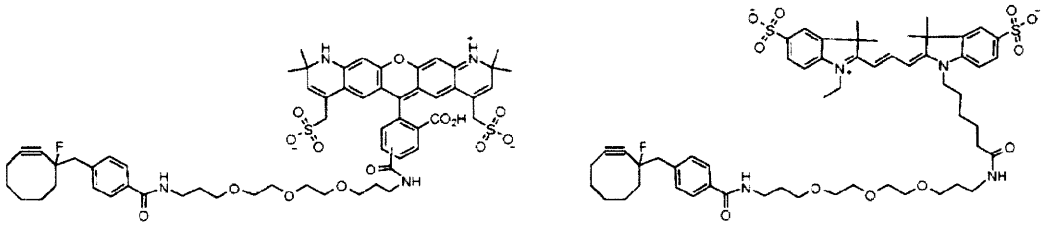


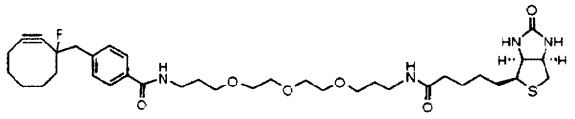
Figure 7-2: Synthetic routes to cyclooctyne-fluorophore conjugates. OCT acid **1** was activated as the pentafluorophenyl (PFP) ester **2**, and then conjugated to a diamino polyethylene glycol (PEG) linker to give OCT-PEG **3**. Reaction of **3** with the activated *N*-hydroxysuccinimidyl (NHS) esters of Cy3 or Alexa Fluor 568 gave the final fluorophore conjugates **4** and **5**.

Figure 7-3 shows the chemical structures of the OCT-PEG-Alexa Fluor 568 (**4**) and OCT-PEG-Cy3 (**5**) conjugates used for cell-surface labeling in this chapter. Additionally, a third probe, an OCT-PEG-biotin (**6**) conjugate was obtained from the Bertozzi lab.



OCT-PEG-Alexa Fluor 568 (**4**)

OCT-PEG-Cy3 (**5**)



OCT-PEG-Biotin (**6**)

Figure 7-3: Chemical structure of OCT-probe membrane impermeant conjugates.

Next, to express the LplA acceptor peptide (LAP) at the surface of mammalian cells, we created an artificial construct by fusing LAP to the N-terminus of cyan fluorescence protein (CFP), which was in turn fused to the extracellular side of the transmembrane (TM) helix of the PDGF receptor (Figure 7-4, top). To perform live cell labeling, HEK cells expressing LAP-CFP-TM were treated first with LplA, azide **7**, and ATP for 1 hour at 32 °C. The introduced azide was then selectively modified with the OCT-fluorophore conjugates by reaction at 25 °C for 20 minutes. Unlike sodium azide, organic azides such as the clinically approved drug AZT[3] are not known to be toxic to cells, but we nevertheless examined the effect of 48-hour exposure to azide **7** on mitochondrial respiration, and found no effect at concentrations less than 750 μM (data not shown).

Figure 7-4 shows the labeling results. We observed specific labeling of transfected (CFP-positive) cells, without labeling of neighboring untransfected cells in the same field of view. We also performed negative controls with azide **7** omitted, or the LAP-CFP-TM construct replaced by its alanine point mutant (alanine at the lysine modification site within LAP), demonstrating the azide-dependent and site-specific nature of the ligation. Interestingly, labeling with Alexa Fluor 568 generated higher background than Cy3 labeling, due to faster non-specific internalization of the probe. We also observed that the extent of non-specific probe internalization was highly cell line-dependent. For example, in contrast to HEK and the cervical carcinoma HeLa cells, which displayed relatively slow endocytosis of the OCT-fluorophores, both Chinese hamster ovary (CHO) cells and African green monkey kidney (COS-7) cells rapidly internalized the OCT-fluorophores, resulting in high background staining, which precluded the cell-surface labeling signal. Recently, Sujiet Puthenveetil, a post-doctoral fellow in our lab, has shown that labeling with OCT-Cy3 can be performed in COS-7 cells if the total labeling time is reduced to ~ 10–15 minutes. This reduction in the labeling time was made possible by the improved ligation kinetics of a different azide analog (an aryl azide derivative) by a mutant of LplA (W37V).

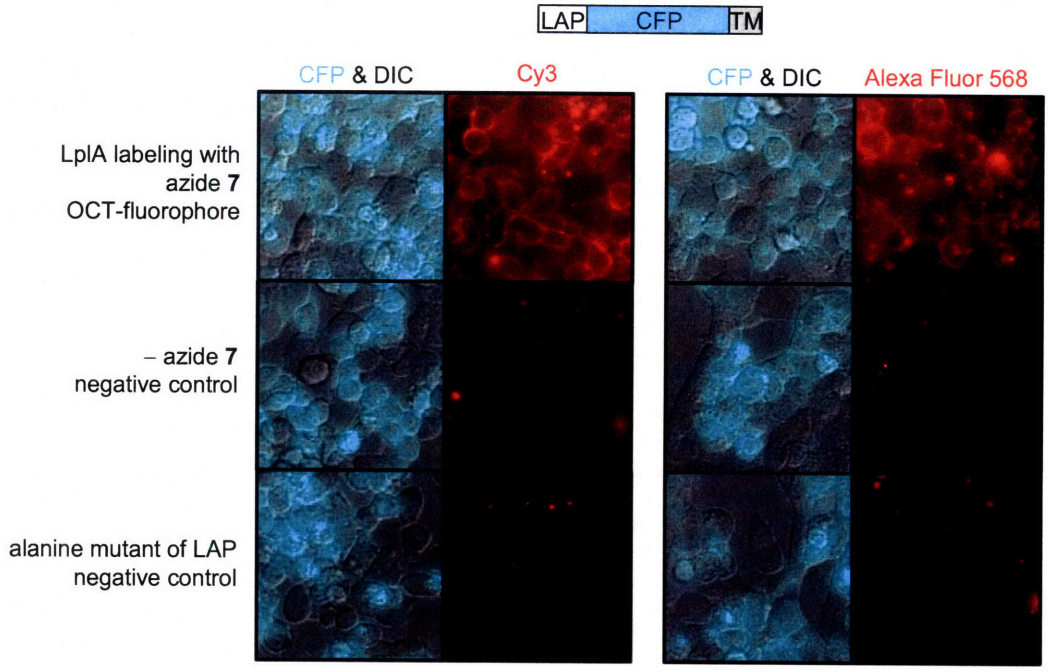


Figure 7-4: Site-specific incorporation of Cy3 and Alexa Fluor 568 fluorophores onto cell-surface proteins in live cells. A LAP-CFP fusion was targeted to the cell surface using a transmembrane (TM) domain. Cell-surface LAP was first labeled with azide **7** by LplA, and the introduced azide was then labeled with a cyclooctyne probe conjugated to Cy3 (left) or Alexa Fluor 568 (right). Live cell images of the introduced fluorophores are shown to the right of the merged CFP and DIC images, which highlight the transfected cells. Negative controls with azide **7** omitted from the labeling reaction, or with the LAP-CFP-TM replaced by its alanine point mutant are shown.

Site-specific labeling of LAP fusion proteins with biotin and quantification of the labeling sensitivity

We also performed labeling with an OCT-biotin conjugate. In this case, labeling was performed in three steps. HEK cells expressing LAP-CFP-TM were treated first with LplA, azide **7**, and ATP for 1 hour at 32 °C, followed by reaction of the azide with OCT-PEG-biotin (**6**) for an additional 15 minutes. Ligated biotin was detected by incubation with streptavidin conjugated to Alexa Fluor 568 for 15 minutes at room temperature. As before, labeling of cell-surface LAP-CFP with the OCT-biotin probe was highly specific (Figure 7-5). Despite the addition of an extra step, the sensitivity of this labeling experiment was higher than that for the Cy3 and Alexa Fluor 568 labeling. We used the wedge method[4] to quantify this sensitivity and determined that cells expressing as little as

5 μM LAP-CFP-TM could be specifically labeled with OCT-biotin, with a signal to background ratio ≥ 3:1. The improved sensitivity may have been a consequence of the multiple copies of Alexa Fluor 568 on each molecule of streptavidin, as well as the absence of background signal from non-specifically internalized biotin probe, which allowed us to increase the temperature of the [3+2] cycloaddition step from 25 °C to 32 °C, and to ultimately reduce the total labeling time for this probe to a minimum of ~ 20 minutes (compared to ~ 1.5 hours for the direct incorporation of Alexa Fluor 568 and Cy3).

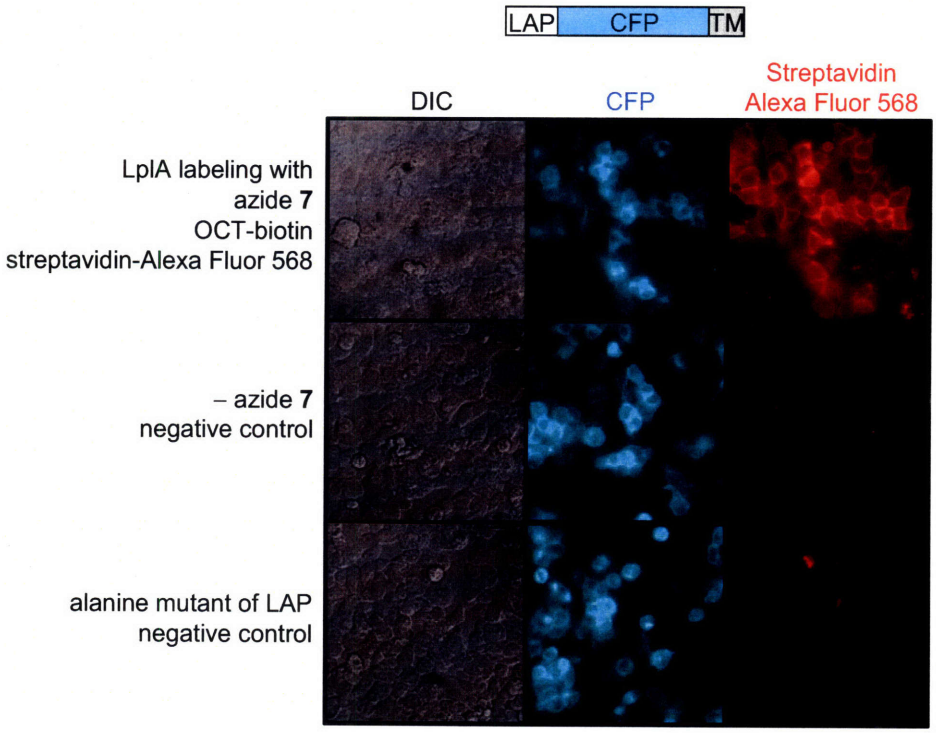


Figure 7-5: Site-specific labeling of LAP-tagged cell-surface proteins with a cyclooctyne-biotin conjugate. A LAP-CFP fusion was targeted to the surface of HEK cells using a transmembrane (TM) domain. Cell-surface LAP was first labeled with LplA and azide **7** followed by OCT-biotin (**6**). Ligated biotin was then detected with streptavidin conjugated to Alexa Fluor 568. Live cell images of the Alexa Fluor 568 fluorophore are shown to the right of the CFP and DIC images, which highlight the transfected cells. Negative controls with azide **7** omitted from the labeling reaction, or with the LAP-CFP-TM replaced by its alanine point mutant are shown.

In terms of sensitivity, the ultimate goal would be to image a real receptor when it is expressed at levels that are physiologically relevant, that is, when it is not highly over expressed. Thus, to further illustrate the sensitivity of LplA labeling for receptor

imaging, we created a LAP fusion to the low-density lipoprotein receptor (LDLR), which functions in the uptake of cholesterol in peripheral tissues of the body[5]. The LAP-LDLR construct was expressed in HEK cells and labeled as above using OCT-biotin. Figure 7-6 shows that when LAP-LDLR levels were similar to endogenous LDLR levels, as measured by staining with LDL conjugated to 1,1'-dioctadecyl-3,3,3',3'-tetramethylindodicarbocyanine (DiD)[6], LplA labeling with OCT-biotin was still clearly detectable.

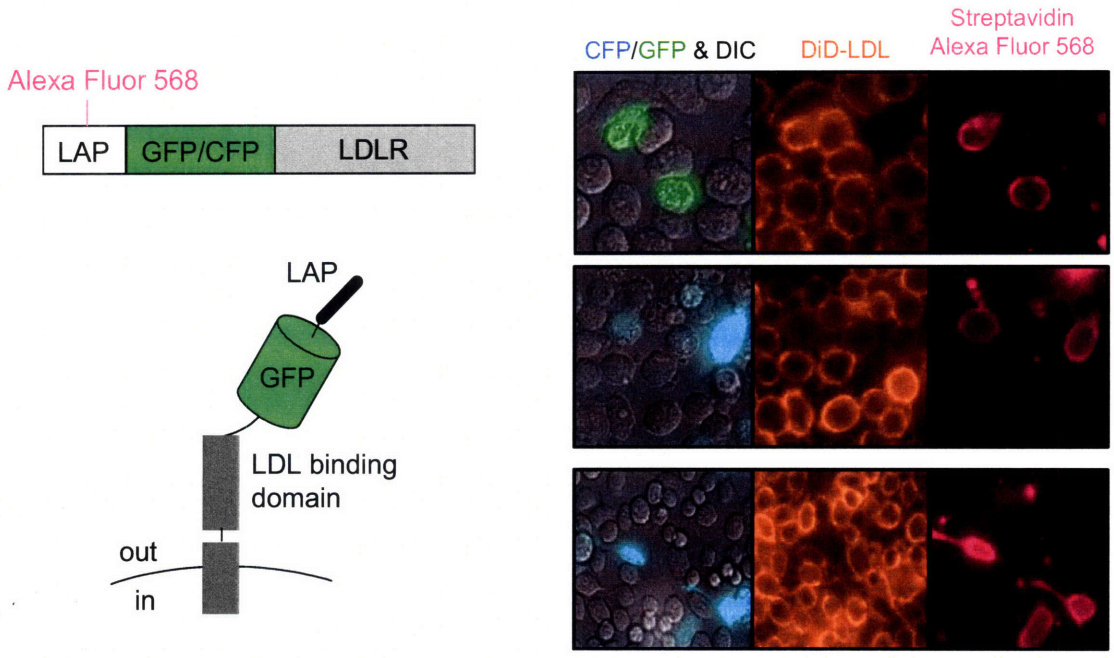


Figure 7-6: LplA detectably labels LAP-tagged low-density lipoprotein receptor expressed at endogenous levels. N-terminally LAP-tagged low density lipoprotein receptor (LDLR) was labeled with LplA and azide **7** followed by OCT-biotin (**6**). Ligated biotin was then detected with streptavidin conjugated to Alexa Fluor 568. Detection of the entire population of LDL receptors (LAP-tagged and endogenous) was achieved using DiD-conjugated LDL. Live cell images of the Alexa Fluor 568 fluorophore are shown to the right of the DiD images. The merged CFP/GFP and DIC images highlight the transfected cells.

Finally, we asked the question of what was the extent of the labeling that resulted in such a high sensitivity. We used a mobility shift assay to quantify the percentage of cell-surface LAP-CFP molecules being modified by azide **7** and OCT-biotin **6**. HEK cells expressing LAP-CFP were first labeled with LplA, azide **7**, and ATP, followed by OCT-biotin. Thereafter, the cells were lysed and the cell extracts were incubated with soluble streptavidin. In-gel fluorescence of the lysates showed that only ~ 3% of the LAP-CFP

molecules were modified with azide **7** and biotin, as visualized by their in-gel mobility shift upon binding to streptavidin (Figure 7-7). The low yield of modification was probably a reflection of the slow ligation of azide **7** to LAP, under conditions in which the concentration of LAP is likely below K_M, combined with the poor kinetics of the [3+2] cycloaddition reaction. Indeed, the Bertozzi group has recently demonstrated that LplA-mediated labeling of cell-surface LAP with a difluorinated-cyclooctyne[7] results in a larger extent of protein modification (J.M. Baskin, personal communication).

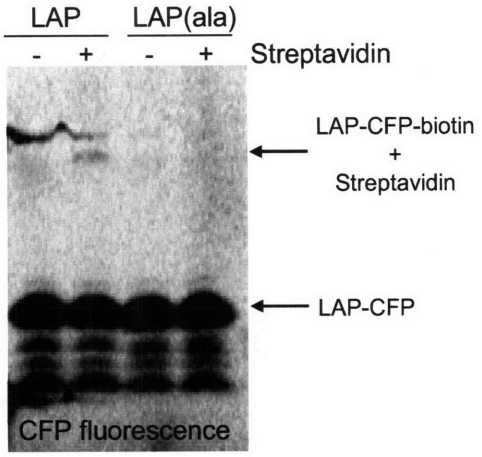


Figure 7-7: Mobility shift assay to quantify cell-surface labeling with LplA and azide 7. HEK cells expressing cell-surface LAP-CFP were labeled with LplA and azide **7** followed by OCT-biotin. After lysis, cell extracts were incubated with soluble streptavidin, and run on SDS-PAGE. CFP fluorescence was visualized in-gel. Negative controls with streptavidin omitted (lanes 1 and 3) or with LAP-CFP replaced by its alanine mutant are shown. Arrows point to unmodified LAP-CFP (below) and labeled LAP-CFP-biotin (above), which displays reduced mobility due to the binding of streptavidin.

LplA labeling is superior to ketone/biotin ligase and transglutaminase labeling in speed, sensitivity, and specificity

Our lab has previously reported two other methods for site-specific labeling of peptide-fused cell surface proteins with small molecule probes. In ketone/biotin ligase labeling, a 15-amino acid "acceptor peptide" (AP) tag is site-specifically conjugated to a ketone analog of biotin by *E. coli* biotin ligase (BirA)[8]. The ketone is then selectively derivatized with a hydrazide or hydroxylamine fluorophore. Transglutaminase labeling uses the guinea pig liver transglutaminase (TGase) to covalently ligate cadaverine-

functionalized fluorophores to proteins fused to a 7-residue glutamine-containing "Q2-tag" recognition sequence for transglutaminase[9].

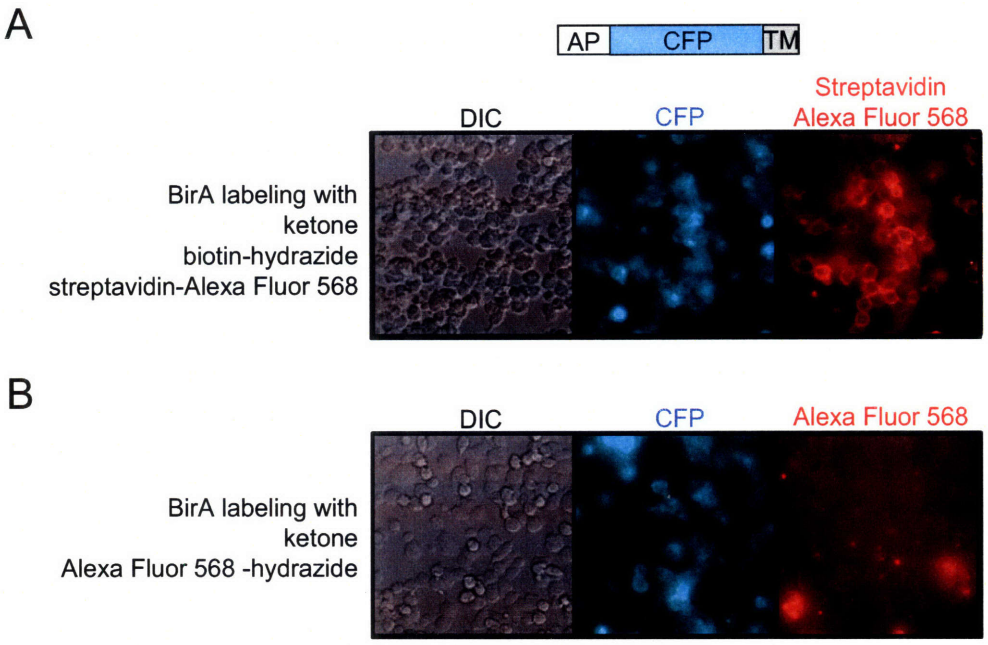


Figure 7-8: Cell-surface labeling of AP-tagged proteins using biotin ligase and a ketone isostere of biotin. An AP-CFP fusion was targeted to the surface of HEK cells using a transmembrane (TM) domain. (**A**) Cell-surface AP was first labeled with *E. coli* biotin ligase (BirA) and ketone followed by biotin-hydrazide. Ligated biotin was then detected with streptavidin conjugated to Alexa Fluor 568. (**B**) Cell-surface AP was first labeled with BirA and ketone followed by Alexa Fluor 568-hydrazide. In both cases (A and B), live cell images of the Alexa Fluor 568 fluorophore are shown to the right of the CFP and DIC images, which highlight the transfected cells.

We directly compared the speed, sensitivity, and specificity of LplA labeling to these methods in side by side labeling experiments. The ketone/biotin ligase labeling was performed on cells expressing an AP-CFP-TM construct (exactly analogous to LAP-CFP-TM) under optimized conditions, using a previously described biotin-benzophenone-hydrazide probe[8]. While the total labeling time for LplA-mediated ligation of OCT-biotin was 45 minutes (Figure 7-5), ketone/biotin ligase labeling required 2 hours and 15 minutes to achieve similar signal to background intensity ratios (Figure 7-8A). Due to the long labeling time, as well as the reduced pH and temperature required for the hydrazone formation step, many cells became unhealthy and rounded. We also tested ketone/biotin ligase labeling for direct incorporation of Alexa Fluor 568 hydrazide

(Figure 7-8B), in a manner analogous to LplA-mediated incorporation of Alexa. Over a wide range of Alexa hydrazide concentrations and different reaction times, the labeling was uniformly unsuccessful. At higher probe concentrations, the non-specific background from Alexa internalization obscured any potential specific cell surface signal. One explanation for the difference in sensitivity is the faster second-order rate constant for the azide-alkyne [3+2] cycloaddition reaction compared to ketone-hydrazide ligation[1,10].

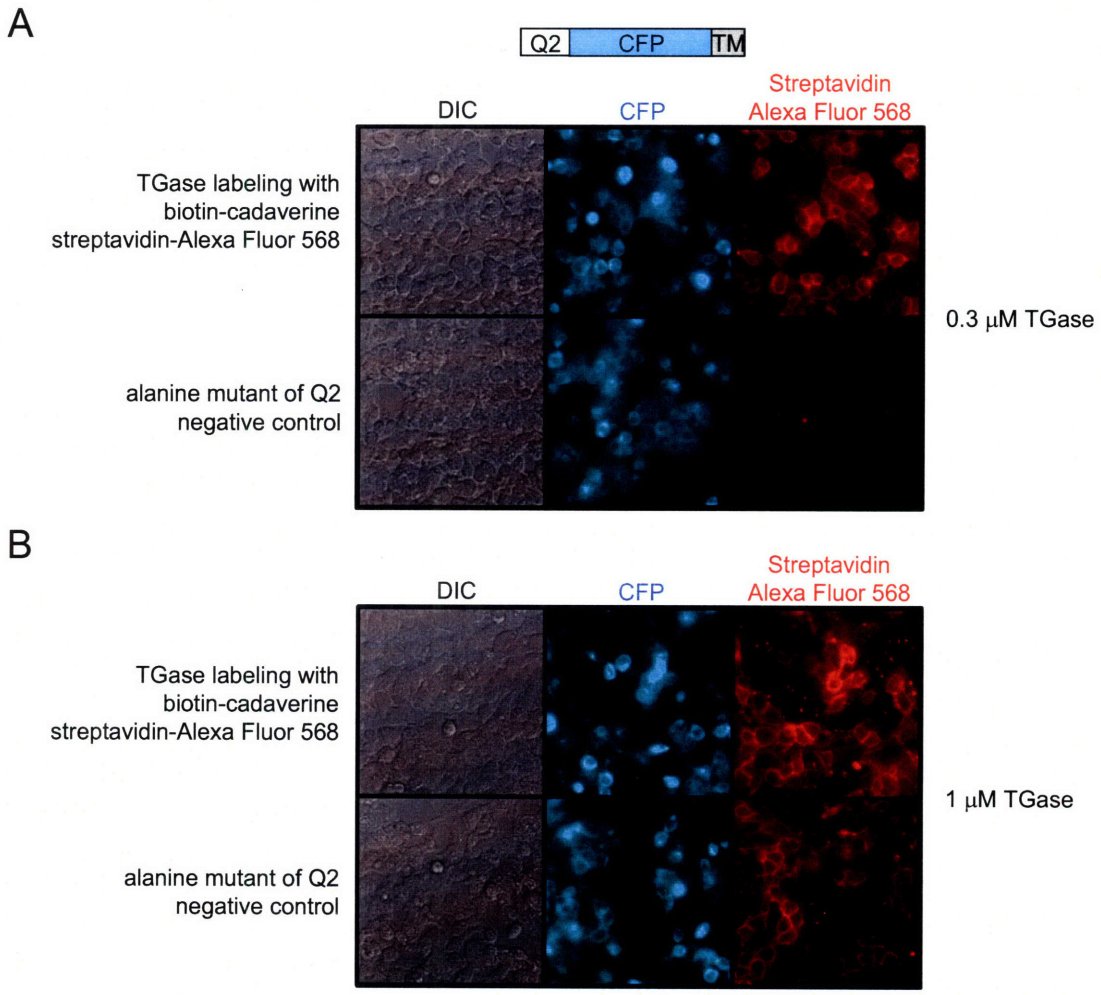


Figure 7-9: Cell-surface labeling of Q2-tagged proteins using transglutaminase. A Q2-CFP fusion was targeted to the surface of HEK cells using a transmembrane (TM) domain. Cell-surface Q2 was labeled with 0.3 μM (**A**) or 1 μM (**B**) transglutaminase (TGase) and biotin-cadaverine. Ligated biotin was then detected with streptavidin conjugated to Alexa Fluor 568. Live cell images of the Alexa Fluor 568 fluorophore are shown to the right of the CFP and DIC images, which highlight the transfected cells.

Transglutaminase labeling was performed on HEK cells expressing the Q2-CFP-TM construct under optimized conditions, using biotin cadaverine, followed by detection with streptavidin-Alexa Fluor 568 conjugate (Figure 7-9). Labeling was comparable to LplA ligation of OCT-biotin, in speed and sensitivity. However, the specificity of transglutaminase labeling was highly variable and difficult to optimize. For example, Figure 7-9A shows good labeling specificity with 0.3 μM transglutaminase, when comparing the Q2-CFP-TM image to the Q2(Ala)-CFP-TM negative control image. However, increasing the enzyme concentration by only 3-fold resulted in the loss of labeling specificity (Figure 7-9B). By contrast, LplA was completely specific for the LAP sequence at enzyme concentrations ranging from 1 μM to 20 μM (data not shown).

LplA labeling is orthogonal to BirA labeling

For many imaging studies, it is desirable to visualize two different receptors at once in the same cell, in order to compare their distribution or trafficking patterns. To develop this capability, we investigated the compatibility of LplA labeling with biotin ligase/streptavidin targeting, a strategy previously developed in our lab by the post-doctoral fellow Mark Howarth[11-14]. Unlike ketone/biotin ligase labeling, our biotin ligase/streptavidin targeting method makes use of site-specific biotin ligation onto AP fusion proteins, followed by recognition with streptavidin conjugates to fluorophores or quantum dots (QDs). While the use of streptavidin increases the total size of the label, the femtomolar affinity of the biotin-streptavidin interaction makes this labeling approach much faster and much more sensitive than ketone/biotin ligase labeling[11].

E. coli LplA and biotin ligase are mechanistically related, and their natural acceptor proteins share some structural and sequence overlap[15]. We showed in Chapter 6, for example, that a peptide derived from a modified biotin carboxyl carrier protein (BCCP) was recognized by LplA *in vitro*. We also showed, however, that a modified AP (with the triad surrounding the lysine site of modification mutated to the lipoylation consensus DKA sequence) was not modified by LplA to any detectable level. Although, the engineered LAP and AP sequences are thus dissimilar, as are the azide **7** and biotin structures, we nonetheless wished to undoubtedly confirm that the two enzyme-mediated

labeling strategies were indeed orthogonal. To test the orthogonality of the two methods, we prepared separate dishes of HEK cells expressing LAP-LDLR with a GFP tag to serve as a marker for transfection, or AP-EGFR (AP fused to the extracellular N-terminus of the EGF receptor[8]) together with a CFP transfection marker. After 16–24 hours of expression, the cells were re-plated together in a single dish. We performed labeling by first adding a mixture of LplA, BirA, azide **7**, biotin, and ATP to the cells. Thereafter, OCT-Cy3 (**5**) was added to derivatize the azide, and streptavidin-QD655 was added to detect the biotin. Figure 7-10 (top) shows that cells expressing LAP-LDLR (as indicated by GFP fluorescence) were selectively labeled with Cy3, while cells expressing AP-EGFR (as indicated by CFP fluorescence) were selectively labeled with QD655. The same results were obtained using LDLR-expressing cells plated beside AP-EphA3-expressing cells (AP fused to the extracellular N-terminus of the receptor for ephrin A3) (Figure 7-10, bottom). Thus, simultaneous labeling of cells with LplA and BirA is possible, without sacrificing the extremely high specificity of each enzyme.

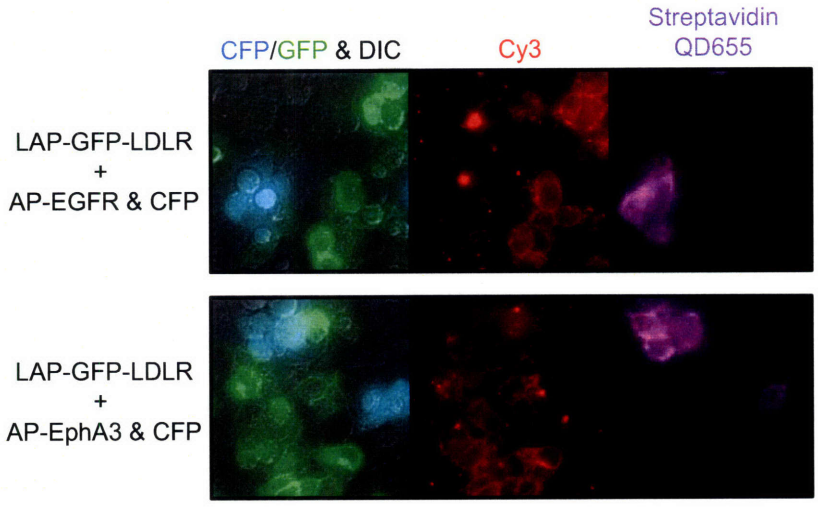


Figure 7-10: Orthogonality test for the simultaneous dual-color labeling of cell-surface proteins with lipoic acid ligase and biotin ligase. Top: Separate dishes of HEK cells were transfected with LAP-GFP-LDLR or AP-EGFR (plus a CFP co-transfection marker). After 24 hours, transfected cells were re-plated together in the same dish. Two-color labeling was performed by treating first with LplA, BirA, azide **7**, and biotin, followed by OCT-Cy3, followed by streptavidin-QD655 (Quantum Dot 655). LAP-GFP-LDLR cells are specifically labeled with Cy3, while AP-EGFR cells (indicated by CFP marker) are specifically labeled with QD655. Bottom: Same, but AP-EGFR was replaced by AP-EphA3.

Two-color receptor labeling on polarized cells using LplA and biotin ligase

We next used the two-color labeling approach to also image LAP- and AP-fused receptors co-expressed in the same cell. EGF receptor and EphA3 are both known to function in cell migration[16,17], and thus we performed imaging on cells migrating toward an artificial wound. HEK cells were co-transfected with either LAP-LDLR and AP-EGFR, or LAP-LDLR and AP-EphA3. After 16–24 hours of expression, the confluent cells were wounded with a pipet tip. We allowed the wound to partially close over 12–18 hours, and then performed simultaneous labeling with Cy3 and Alexa Fluor 488 conjugated to monovalent streptavidin. Monovalent streptavidin was engineered by Mark Howarth in our lab, and it has a single femtomolar-affinity biotin binding site[12]. Monovalent streptavidin is preferred for receptor imaging over the tetravalent wild-type protein (four biotin binding sites) because it eliminates artifacts introduced by streptavidin-mediated receptor crosslinking. Figure 7-11 shows the results of the two-color imaging experiment. We observed that Cy3-labeled LDLR was evenly distributed on the surface of the HEK cells, whereas Alexa Fluor 488-labeled EGFR and EphA3 were both asymmetrically concentrated at the leading edge of the polarized cells. The same patterns were also observed when the LAP and AP tags were swapped (AP-LDLR and LAP-EGFR), suggesting that the localization patterns do not reflect artifacts of AP and LAP labeling (data not shown). It must be noted however, that the signal obtained for LplA-mediated labeling of LAP-EGFR with OCT-Cy3 was much lower than the LAP-LDLR signal. We think that this is a reflection of the relative expression levels of both proteins at the cell-surface, since we observed the same trend for BirA-mediated labeling of the corresponding AP-tagged receptors (Mark Howarth, personal communication).

While the polarization of AP-EGFR to the leading edge of migrating cells was an expected result, and it had previously been observed using antibody detection on epithelial cells[18], the pattern of AP-EphA3 staining was somewhat surprising. Previous work had shown that *trans* interactions between EphA3 and ephrin ligand expressed on the surface of contacting cells play a role in developmental cell migration[19] and tumor invasion[20]. However, it was unclear that un-liganded EphA3 should function in migratory processes. Our observation of EphA3 accumulation at the free, leading edge of polarized

cells suggested that unactivated EphA3 may play a role in cell signaling, or that EphA3 may be constitutively linked to the actin cytoskeleton.

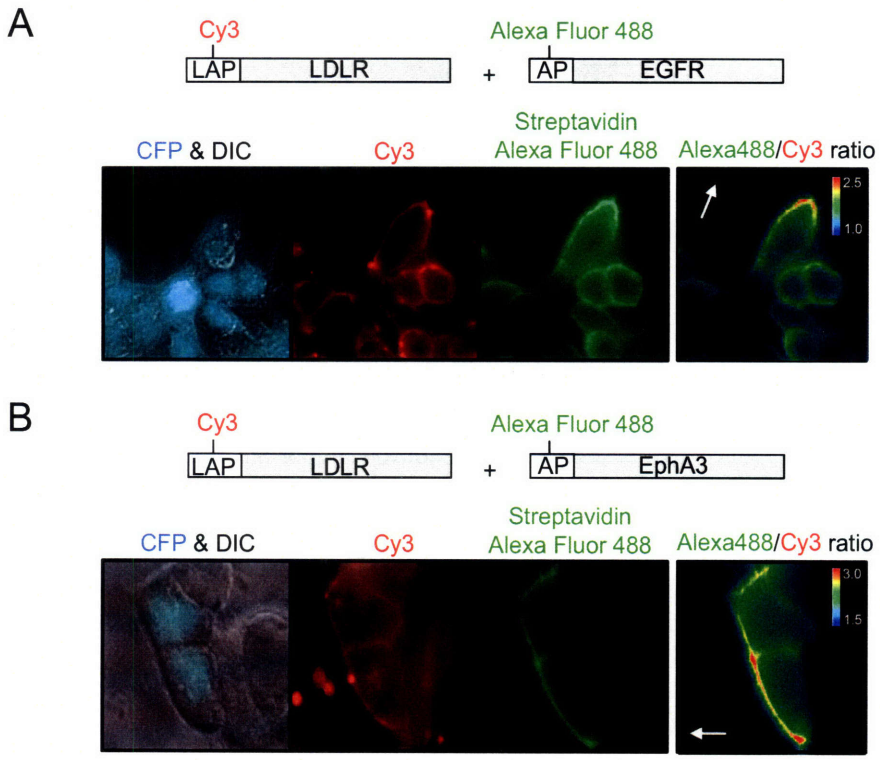


Figure 7-11: Simultaneous labeling and imaging of two receptors in polarized cells in a wound healing assay. HEK cells co-expressing a LAP-LDLR fusion and either AP-EGFR (**A**) or AP-EphA3 (**B**) were labeled during wound healing by first treating with LplA, BirA, azide **7**, and biotin, followed by OCT-Cy3 to derivatize the azide, followed by monovalent streptavidin-Alexa Fluor 488[12] to detect the biotin. The Cy3 images show the non-polarized distribution of surface LAP-LDLR. The Alexa Fluor 488 images show the polarized distribution of AP-EGFR (A) and AP-EphA3 (B) at the wound edge. CFP is a transfection marker. The images on the far right depict the intensity ratios of Alexa Fluor 488 and Cy3. The white arrows point toward the wound.

One-step cell-surface labeling with LplA using a coumarin probe

In a separate effort, Dr. Hemanta Baruah and others in our laboratory have explored the substrate tolerance of several LplA point mutants. As previously mentioned in Chapter 6, direct incorporation of a fluorophore or a photoaffinity probe is desirable because it could in principle shorten the labeling time compared to the use of functional group handles followed by bioorhotogonal ligation.

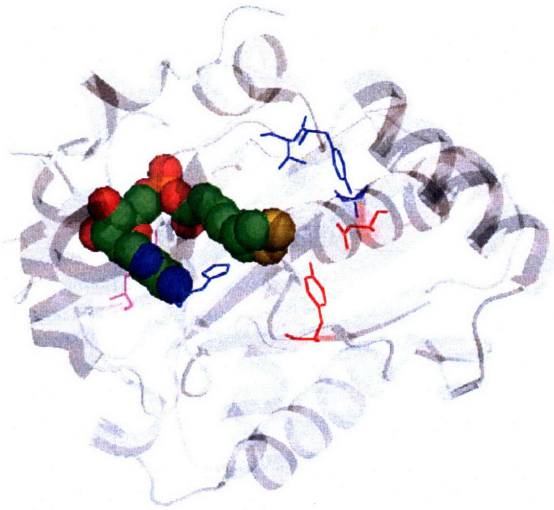


Figure 7-12: LplA gatekeeper residues mutated to allow incorporation of larger analogs of lipoic acid. View of the active site of *Thermoplasma acidophilum* around the bound lipoyl-AMP (PDB: 2ART)[21]. Lipoyl-AMP is shown in CPK format. Important amino acid residues are shown in ball and stick format. The catalytic Lysine 145 is colored magenta. Residues colored red (Asp21 and Tyr39) are the analogous residues to those mutated in *E. coli* that allowed incorporation of coumarin derivatives (*E. coli* LplA Glu20 and Trp37). Residues colored blue (Tyr73, Thr74, His80, and Leu158) are the analogous residues to those mutated in *E. coli* that allowed incorporation of other probes, such as an aryl azide photoaffinity probe (*E. coli* LplA Ser71, Ser72, His79, and Phe147).

As expected, none of the most interesting probes were incorporated by the wild-type enzyme. Based on the published crystal structure of *Thermoplasma acidophilum* LplA in complex with lipoyl-AMP[21,22], Dr. Hemanta Baruah and Prof. Alice Ting rationally designed a panel of LplA mutants, which were then tested by Dr. Baruah for incorporation of several probes. Figure 7-12 highlights those residues that have proven important as active site gatekeepers, and whose mutation resulted in the incorporation of one or more probes. The main objective for the tested mutations was to enlarge the active site cavity that encloses the dithiolane ring, in an attempt to accommodate larger groups. Of particular interest was the tryptophan 37 residue (Tyrosine 39 in Ta LplA), which appears in the Ta crystal structure at ~ 4 Å from one of the sulfur atoms in the dithiolane ring. Mutation of Trp37 to a hydrophobic valine or isoleucine allowed incorporation of a coumarin fluorophore (coumarins **1** and **2**, Figure 7-14A), whereas mutation to a polar residue such as serine resulted in the highly efficient ligation of an aryl azide photoaffinity probe (Figure 7-14A). Remarkably, Dr. Baruah has now shown that ligation of the aryl azide derivative by LplA-W37S onto LAP is catalytically more efficient than

the ligation of lipoic acid by the wild-type enzyme (Baruah et al., *Angew. Chem. Int. Ed Engl.*, in press).

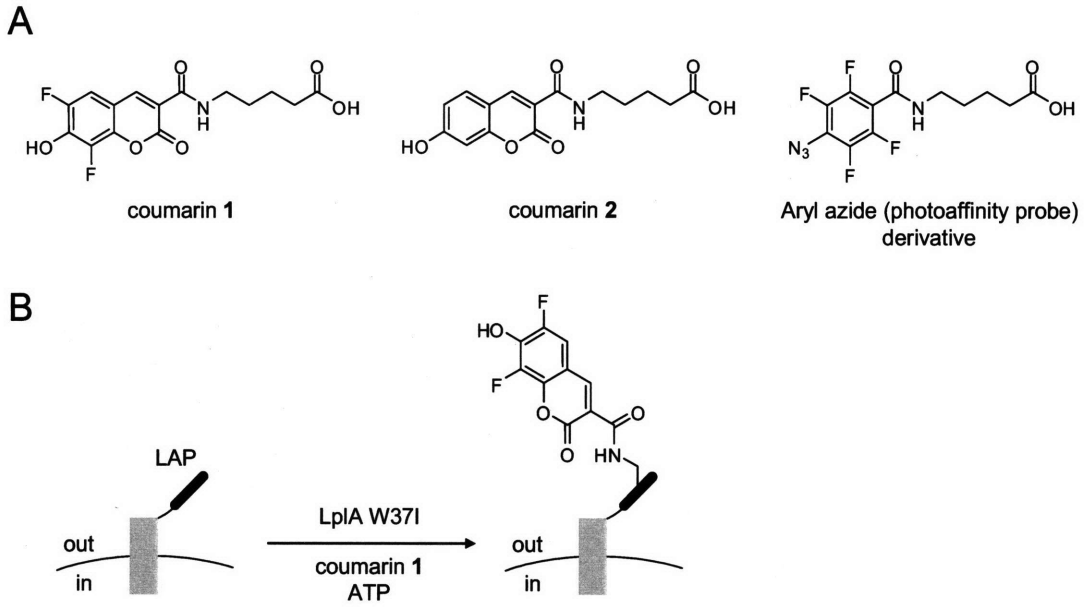


Figure 7-13: LplA-catalyzed ligation of coumarin fluorophores. (**A**) Structure of coumarin probes and the aryl azide photoaffinity probe incorporated by the W37I and W37V mutants of LplA. (**B**) One-step cell-surface labeling of LAP-tagged proteins. Cell-surface LAP is labeled in a single step by an LplA mutant (e.g., W37I) using a coumarin derivative (e.g., coumarin **1**) and ATP.

We thus decided to test the direct incorporation of coumarins **1** and **2** by W37I on the surface of HEK cells. Unfortunately, the W37I mutant does not recognize the LAP as efficiently as WT LplA. Therefore, we created an artificial construct by fusing the hybrid lipoyl domain derived from E2p[23] to the N-terminus of cyan fluorescence protein (CFP), which was in turn fused to the extracellular side of the transmembrane (TM) helix of the PDGF receptor, to give E2p-CFP-TM; a construct analogous to the previously described LAP-CFP-TM. Because the excitation and emission spectra of coumarin and CFP overlap, we mutated Trp66 in CFP to leucine to abolish its fluorescence (E2p-CFP*-TM, Figure 7-14, top). To perform one-step live cell labeling, HEK cells expressing E2p-CFP*-TM were treated with LplA, the coumarin probe, and ATP for 10 minutes at 32 °C (Figure 7-13B). Even with such a short labeling time, we observed robust and specific labeling of HEK cells expressing cell-surface E2p (Figure 7-14). Both, coumarin **1** and its non-fluorinated version coumarin **2,** resulted in a labeling signal that was ~ 3-fold higher

than that obtained using azide **7** and OCT-Cy3. Although the signal observed with coumarin **2** may seem dimmer, this was a mere reflection of its higher pKa (~ 7.5), which means that at the imaging pH of 7.4, approximately half of the molecules were in a non-fluorescent state. The lower signal could also be caused by faster photobleaching of coumarin **2**. A combined labeling with coumarin **1** and the pH-sensitive coumarin **2** should in principle allow ratiometric studies of receptor endocytosis.

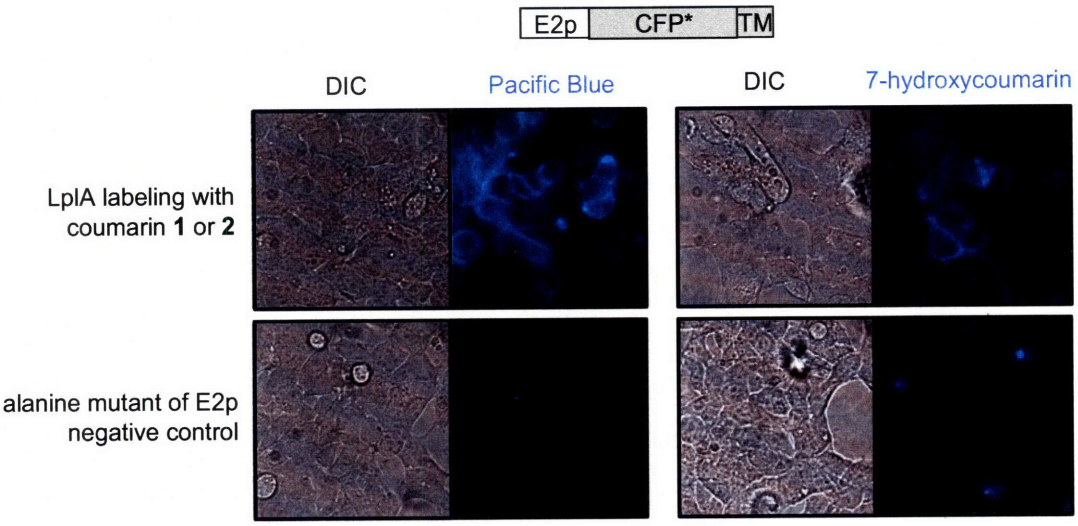


Figure 7-14: Direct fluorophore incorporation onto cell-surface proteins in live cells. An E2p fusion to a modified non-fluorescent CFP was targeted to the surface of HEK cells using a transmembrane (TM) domain. Cell-surface E2p was labeled with LplA W37I and either coumarin **1** (Pacific Blue, left) or coumarin **2** (7-hydroxycoumarin, right). Live cell images of the coumarin fluorophore are shown to the right DIC images. A negative control with the E2p-CFP-TM replaced by its alanine point mutant is shown.

Conclusions

In summary, we have developed a new methodology for labeling cell surface proteins fused to a 22-amino acid recognition sequence for *E. coli* LplA. Small, non-crosslinking probes such as Cy3 and Alexa Fluor were site-specifically and covalently conjugated to the LAP peptide in about 1 hour, whereas biotin could be ligated in as little as 20 minutes. An important feature of this methodology is its generality; any cell surface protein in any cell type can be labeled with any chemical moiety that can be

functionalized with a cyclooctyne. Direct comparison of LplA labeling to our two previous methods showed that LplA was superior to ketone/biotin ligase tagging in speed and sensitivity, and far superior to transglutaminase labeling in specificity over a wide range of conditions.

We used LplA methodology in combination with biotin ligase labeling to image two different receptors in the same cell. We found that the EGF receptor and EphA3 both localized to the leading edge of fibroblasts migrating towards a wound. Many problems in receptor biology would benefit from simultaneous imaging of two or more different proteins at once in the same living cell, instead of separate experiments involving one-color labeling of each receptor. The combination of LplA and BirA tagging, which can be performed simultaneously due to the orthogonality of the labeling reaction components, should provide access to such experiments.

Finally, we have also shown that direct fluorophore ligation is possible, and can in fact accelerate the reaction to result in a total labeling time of 10 minutes. This procedure should allow studies of faster biological processes and is better suited to imaging the trafficking of cell surface receptors.

Experimental

General synthetic methods

Reagents were purchased from Sigma-Aldrich, Alfa Aesar, TCI America, Invitrogen, or GE Healthcare and used without further purification. Analytical thin layer chromatography was performed using 0.25 mm silica gel 60 F_{254} plates and visualized with ninhydrin or bromocresol. Flash column chromatography was carried out using silica gel (ICN SiliTech 32-63D). Mass spectra were recorded on an Applied Biosystems 200 QTRAP Mass Spectrometer using electrospray ionization. HPLC was performed on a Varian Prostar Instrument equipped with an autosampler and photo-diode-array detector. For analytical HPLC, a reverse-phase 250 × 4.6 mm Microsorb-MV 300 C18 column was

used. For preparative HPLC, a reverse-phase 250 × 10 mm Microsorb-MV 100 C18 column was used. Chromatograms were recorded at 210 nm unless otherwise noted.

Synthesis of OCT-fluorophore conjugates

Synthesis of OCT-PEG (3)

To a solution of OCT acid **1**[1] (32 mg, 123 μmol) in dry dichloromethane (DCM, 750 μL) was added triethylamine (TEA, 34 μL, 246 μmol). The mixture was stirred for 5 minutes at ambient temperature. Pentafluorophenyl trifluoroacetate (PFP-TFA, 42 μL, 246 μmol) was added slowly to the reaction mixture over 3 minutes and the reaction was allowed to proceed for 3 hours. The reaction mixture was concentrated in *vacuo*, then purified by silica chromatography (10% ethyl acetate in hexane) to afford OCT-PFP **2** as a colorless solid (45.6 mg, 107 μmol, 87.3%). The solid was immediately dissolved in dry DCM (1 mL), and combined with TEA (29 μL, 214 μmol). *O,O'*-Bis(3-aminopropyl)diethylene glycol (diamino-PEG, 71 μL, 535 μmol) was added, and the reaction mixture was allowed to stir overnight at 21 °C. The crude mixture was purified on a silica flash column (5:3:1 ethyl acetate : methanol : water with 4.3% triethylamine). The purified product was concentrated under reduced pressure to yield OCT-PEG **3** (27.8 mg, 60 μmol, 56%). ESI-MS calculated for $[M + H]^+$: 463.29; observed 463.44.

Synthesis of OCT-PEG-Alexa Fluor 568 (4) and OCT-PEG-Cy3 (5)

To a solution of OCT-PEG **3** (0.6 mg, 1.3 μmol) in anhydrous DMSO (200 μL) was added TEA (0.6 μL, 4 μmol). The *N*-hydroxysuccinimidyl ester of either Cy3 or Alexa Fluor 568 (1.3 μmol) was then added and the reaction was allowed to stir for 4 hours at 21 °C. The crude product was purified by HPLC on a reverse phase column. Chromatograms were recorded both at 210 nm and 550 nm. The following conditions were used for elution: 30–80% acetonitrile in water over 30 minutes; flow rate 5.0 mL/minute. Solvent was removed in *vacuo* to afford the OCT-PEG-fluorophore conjugates. Estimated yield 50–65%. ESI-MS for OCT-PEG-Alexa Fluor 568 (**4**): calculated for $[M - H]^-$: 1137.42; observed 1137.54. ESI-MS for OCT-PEG-Cy3 (**5**):

calculated for $[M + H]^+$: 1075.48; observed 1075.08. In addition, the presence of an intact OCT moiety was confirmed by reaction of 100 μM OCT-PEG-Alexa Fluor 568 (**4**) or OCT-PEG-Cy3 (**5**) with 1mM azide **7** in DMSO at 30 °C for 6 hours and detection of the triazole cycloadduct product by MS. ESI-MS for OCT-PEG-Alexa Fluor 568-cycloadduct: calculated for $[M - H]^-$: 1322.54; observed 1322.22. ESI-MS for OCT-PEG-Cy3-cycloadduct: calculated for $[M - H]^-$: 1258.60; observed 1258.56.

Synthesis of coumarin probes

Synthesis of coumarin 1

To a solution of 5-aminovaleric acid (1.6 mg, 14 μmol) in anhydrous DMSO (100 μL) was added the *N*-hydroxysuccinimidyl ester of 3-carboxy-6,8-difluoro-7-hydroxycoumarin (2.5 mg, 7 μmol) and the reaction mixture was stirred overnight at ambient temperature in the dark. Excess insoluble 5-aminovaleric acid was discarded by centrifugation. The product was used without further purification. ESI-MS for coumarin **1** calculated for $[M - H]^-$: 340.07; observed 340.26.

Synthesis of coumarin 2

To a solution of 5-aminovaleric acid (65.5 mg, 560 μmol) in anhydrous DMSO (400 μL) was added the *N*-hydroxysuccinimidyl ester of 3-carboxy-7-hydroxycoumarin (34 mg, 112 μmol) and the reaction mixture was stirred overnight at ambient temperature in the dark. Excess insoluble 5-aminovaleric acid was discarded by centrifugation. The product was used without further purification. ESI-MS for coumarin **2** calculated for $[M - H]^-$: 304.09; observed 304.26.

Cloning of LAP-CFP-TM for mammalian surface expression

The LAP peptide was inserted between the *BglII* and *AscI* sites of the AP-CFP-TM plasmid[8]. Primers 5′ AAAAA *AGA TCT* GGC GGC GAC GAA GTA CTG GTT GAA ATC GAA ACC GAC AAA GCA GTT CTG GAA GTA CCG GGC GGT GAG GAG GAG *GGC GCG CCA* AAAA and its reverse complement were annealed together, digested with *BglII* and *AscI*,

and ligated in-frame to *BglII/AscI* digested AP-CFP-TM. To create the LAP(Ala) mutant, we performed QuikChange using the primer 5′ G GTT GAA ATC GAA ACC GAC GCC GCA GTT CTG GAA GTA CCG G and its reverse complement.

Cloning of LAP-LDLR and LAP-GFP-LDLR

To clone LAP-GFP-LDLR, the LAP sequence was inserted in the leader sequence of LDLR, between Ser27 and Thr28, using the inverse PCR method described by Gama and Breitwieser[24] with the primers 5′ CG GGC GGT GAG GAG GAG ACT GTG AGC AAG GGC GAG GAG and 5′ GTA CCT CCA GAA CTG CTT TGT CGG TTT CGA TTT CAA CCA GTA CTT CGT CAC TTC TGT CGC CAA CTG CAG. The PCR template was the pEGFP-LDLR plasmid, a gift from Tom Kirchhausen. To clone LAP-LDLR, the GFP gene was deleted by QuikChange using primers 5′ GGA GGT ACC GGC ATC AGC AGA CGG CGG GGG AGA ATT CGA CAG ATG TG and its reverse complement.

Cloning of E2p-CFP*-TM for mammalian surface expression

The gene encoding for a 9kD hybrid lipoyl domain derived from the second subunit of the *E. coli* pyruvate dehydrogenase[23] was inserted between the *BglII* and *AscI* sites of the AP-CFP-TM plasmid[8]. The gene was PCR-amplified from the original pGS331 plasmid[23], a gift from John Cronan, using primers 5′ AAAA AGA TCT GGC GGC ATG GCT ATC GAA ATC AAA GTA CCG G and 5′ TTTT GG CGC GCC CGC AGG AGC TGC CGC AGG CG. The resulting PCR product was digested with *BglII* and *AscI* and ligated in-frame to *BglII/AscI* digested AP-CFP-TM, to give E2p-CFP-TM. To create the non-fluorescent mutant E2p-CFP*-TM, we performed QuikChange using the primer 5′ G ACC ACC CTG ACC CTG GGC GTG CAG TGC TTC and its reverse complement. To create the E2p(Ala) mutant, we performed QuikChange using the primer 5′ C ACC GTA GAA GGC GAC GCA GCT TCT ATG GAA GTT CCG and its reverse complement.

Live cell labeling of cell surface LAP-CFP-TM with OCT-fluorophore conjugates

HEK 293T cells were transfected with the LAP-CFP-TM expression plasmid using Lipofectamine 2000. After 36–48 hours at 37 °C, the cells were washed twice with fresh growth media (Dulbecco's Modified Eagle's Medium with 10% fetal bovine serum and 1% penicillin/streptomycin). Enzymatic ligation of azide **7** was performed in complete growth media with 10 μM LplA, 350 μM azide **7**, 1 mM ATP, and 5 mM magnesium acetate for 60 minutes at 32 °C. Cells were rinsed three times with growth media, and incubated for 20 minutes at 21 °C with 200–400 μM OCT-Cy3 or 100–200 μM OCT-Alexa Fluor 568. Thereafter, the cells were washed once with growth media, twice with a 1% bovine serum albumin (BSA) solution in Dulbecco's Phosphate-Buffered Saline (DPBS) pH 7.4, and twice more with DPBS alone. Labeled cells were imaged in the same buffer on a Zeiss Axiovert 200M inverted epifluorescence microscope using a 40x oil-immersion lens. CFP (420/20 excitation, 450 dichroic, 475/40 emission), Cy3 and AlexaFluor568 (560/20 excitation, 585 dichroic, 605/30) and differential interference contrast (DIC) images (630/10 emission) were collected and analyzed using Slidebook software (Intelligent Imaging Innovations). Fluorescence images were normalized to the same intensity range. Acquisition times ranged from 10–250 milliseconds.

Live cell labeling of cell surface LAP-CFP-TM with OCT-biotin

HEK 293T cells were transfected with the LAP-CFP-TM plasmid using Lipofectamine 2000. After 36–48 hours at 37 °C, the cells were washed twice with fresh growth media (DMEM supplemented with 10% FBS and 1% penicillin/streptomycin). Enzymatic ligation of azide **7** was performed in complete growth media with 10 μM LplA, 350 μM azide **7**, 1 mM ATP, and 5 mM magnesium acetate for 15 minutes at 32 °C. Cells were then rinsed three times with growth media, and incubated for 15 minutes at 32 °C with 250 μM OCT-biotin. Thereafter, the cells were washed twice with growth media and incubated with 30 μg/ml streptavidin-Alexa Fluor 568 (Invitrogen) for 15 minutes at 21 °C. The cells were washed once with growth media at

21 °C and twice with ice-cold DPBS, pH 7.4, and imaged in the same buffer on a Zeiss Axiovert 200M as described above.

Determination of LplA labeling sensitivity with LAP-CFP-TM using the wedge method

We used the wedge method to estimate the concentration of CFP in single cells expressing the LAP-CFP-TM construct[4,8]. A wedge-shaped microchamber was constructed from three glass cover slips. The length along the x direction was 6.5 mm and the height of the chamber (z-direction) increased linearly from 0 to 150 μm. The chamber was filled with a solution of 10 μM purified CFP in PBS pH 7.4. The fluorescence of the wedge was imaged under conditions identical to those used for cellular imaging. We assumed an average cell thickness of 5 μm and therefore interpolated to the region of the wedge with thickness 5 μm and used the fluorescence intensity measured there as a reference standard for CFP concentration measured in single cells. Using the wedge for comparison, we imaged two samples of HEK and HeLa cells expressing LAP-CFP-TM. Examination of the CFP channel images for cells that display clear OCT-biotin labeling (signal to background ≥ 3:1) showed that the CFP concentrations ranged from 5 μM to >1 mM. We therefore concluded that cells expressing as little as 5 μM LAP-CFP-TM could be labeled by our method. This represents an upper limit to the labeling sensitivity of our methodology.

Live cell labeling of LAP-LDLR with OCT-biotin and DiD-LDL

HEK 293T cells were either singly transfected with LAP-GFP-LDLR or co-transfected with a mixture of LAP-LDLR and CFP-pcDNA3 in a 6:1 ratio using Lipofectamine 2000. After 36–48 hours at 37 °C, cells were washed twice with fresh growth media (DMEM supplemented with 10% FBS and 1% penicillin/streptomycin). Enzymatic ligation of azide **7** was performed in complete growth media with 10 μM LplA, 350 μM azide **7**, 1 mM ATP, and 5 mM magnesium acetate for 20 minutes at 32 °C. Cells were then rinsed three times with growth media, and incubated for 30 minutes at 32 °C with 250 μM OCT-biotin. Thereafter, the cells were washed twice

with growth media and incubated with 30 μg/ml streptavidin-Alexa Fluor 568 (Invitrogen) and DiD-LDL (1 μg/ml)[6] for 15 minutes at 4 °C. The cells were washed once with growth media at 4 °C and twice with ice-cold DPBS, pH 7.4, and imaged in the same buffer as described above. DiD was imaged using 630DF10 excitation, 660DRLP dichroic, and 680DF30 emission. Acquisition times ranged from 10–200 ms.

Determination of LplA labeling sensitivity with LAP-CFP-TM using a mobility shift assay

HEK 293T cells were transfected and labeled with azide **7** and OCT-biotin as above. After incubation with OCT-biotin, cells were washed twice with DPBS and lysed. Lysates were generated using a RIPA lysis buffer to efficiently extract membrane proteins, as follows. Cells were lifted from the plates, concentrated by centrifugation, and resuspended in 25 mM Tris-HCl pH 7.6, 150 mM NaCl, 1% Triton X-100, 1% sodium deoxycholate, 0.1 % sodium dodecyl sulfate (SDS), 1 mM phenylmethylsulphonyl fluoride (PMSF), and protease inhibitor cocktail (Calbiochem). After incubation at 4 °C for 30 minutes, the crude lysates were clarified by centrifugation. 10 μM streptavidin was then added to the lysates and incubated at room temperature for 1 hour. The reaction mixtures were analyzed by 12% SDS-PAGE without boiling the samples, in order to prevent unfolding of CFP and to preserve the biotin-streptavidin binding. In-gel fluorescence was visualized on a Storm 860 instrument.

Live cell labeling of cell surface AP-CFP-TM with BirA and ketone

HEK 293T cells were transfected with AP-CFP-TM plasmid[8] using Lipofectamine 2000. After 36–48 hours at 37 °C, the cells were washed twice with DPBS pH 7.4 and labeling was performed as previously reported[8]. Briefly, enzymatic ligation of ketone to AP-CFP-TM was performed in DPBS pH 7.4, with 0.2 μM BirA, 1 mM ketone **1**[8], 1 mM ATP, and 5 mM magnesium acetate for 60 minutes at 32 °C. Cells were then washed twice with DPBS, pH 6.2 and incubated for 60 minutes at 16 °C (to reduce endocytosis) with 1 mM benzophenone-biotin-hydrazide[8] in DPBS pH 6.2. Thereafter, the cells were washed twice with ice-cold DPBS pH 7.4, and incubated with 30 μg/mL

streptavidin-Alexa Fluor 568 (Invitrogen) in DPBS pH 7.4 and 1% BSA for 15 minutes at 4 °C. The cells were washed twice with DPBS pH 7.4, and imaged in the same buffer as described above.

Live cell labeling of cell surface Q2-CFP-TM with transglutaminase

HEK 293T cells were transfected with Q2-CFP-TM plasmid[9] using Lipofectamine 2000. After 36–48 hours at 37 °C, the cells were washed twice with DMEM, and labeling was performed as previously reported[9]. Briefly, enzymatic ligation of biotin cadaverine to Q2-CFP-TM was performed in DMEM, with 0.3 μM or 1 μM guinea pig liver transglutaminase (TGase, NZyme BioTec GmbH), 0.5 mM biotin cadaverine, and 12 mM $CaCl_2$ for 30 minutes at 37 °C. Cells were then washed twice with DPBS pH 7.4, and incubated for 15 minutes at 4 °C (to reduce endocytosis) with 30 μg/mL streptavidin-Alexa Fluor 568 (Invitrogen) in DPBS pH 7.4, and 1% BSA. The cells were washed three times with ice-cold DPBS pH 7.4, and imaged in the same buffer as described above.

Orthogonality test for two-color labeling with LplA and BirA

HEK 293T cells were either singly transfected with LAP-GFP-LDLR or co-transfected with a mixture of AP-EGFR (or AP-EphA3) and CFP-pcDNA3 in a 6:1 ratio using Lipofectamine 2000. 24 hours after transfection, the cells were re-plated together in a 1:1 ratio. After further incubation for 24 hours at 37 °C, the cells were washed twice with fresh growth media (DMEM with 10% FBS and 1% penicillin/streptomycin). Simultaneous enzymatic ligations of azide **7** and biotin were performed in complete growth media with 5 μM BirA, 10 μM LplA, 50 μM biotin, 350 μM azide **7**, 1 mM ATP, and 5 mM magnesium acetate for 60 minutes at 32 °C. Cells were then rinsed three times with growth media, and incubated for 20 minutes at 21 °C with 200–400 μM OCT-Cy3. Biotin was detected with streptavidin-QD655 (13 nM, Invitrogen) for 10 minutes at 4 °C. Thereafter, the cells were washed once with ice-cold 1% BSA in DPBS, pH 7.4 and twice more with ice-cold DPBS, pH 7.4. Labeled cells were imaged as described above. The

GFP filter set was 495/20 excitation, 515 dichroic, 530/30 emission; the QD655 filter set was 405/20, 585 dichroic, and 655/20 emission.

Two-color live cell labeling with LplA and biotin ligase

HEK 239T cells were co-transfected with the LAP-LDLR and AP-EGFR[16] plasmids in a 1:2 ratio, or with the LAP-LDLR and AP-EphA3 (a gift from M. Lackmann, Monash University) plasmids in a 2:1 ratio. 24 hours after transfection, the cells were wounded with a pipet tip and allowed to heal over 16–24 hours. For labeling, cells were washed twice with complete growth media, and then incubated with 5 μM BirA, 10 μM LplA, 50 μM biotin, 350 μM azide **7**, 1 mM ATP, and 5 mM magnesium acetate for 60 minutes at 32 °C. Cells were then rinsed three times with growth media, and incubated for 20 minutes at 21 °C with 200–400 μM OCT-Cy3. Biotin was detected by staining with 50 μg/mL monovalent streptavidin-Alexa Fluor 488 (prepared as previously described[12]) for 10 minutes at 4 °C. The cells were washed once with ice-cold 1% BSA in DPBS pH 7.4, then twice with ice-cold DPBS, before imaging in the same buffer using the configuration described above. The Alexa Fluor 488 filter set was 495/20 excitation, 515 dichroic, 530/30 emission.

One-step live cell labeling of cell surface LAP-CFP-TM with coumarin fluorophores

HEK 293T cells were transfected with the E2p-CFP*-TM expression plasmid using Lipofectamine 2000. After 36–48 hours at 37 °C, the cells were washed twice with fresh growth media (Dulbecco's Modified Eagle's Medium with 10% fetal bovine serum and 1% penicillin/streptomycin). Enzymatic ligation of the coumarin fluorophore was performed in complete growth media with 3 μM LplA W37I, 340 μM coumarin **1** or coumarin **2**, 1 mM ATP, and 5 mM magnesium acetate for 10–30 minutes at 32 °C. Thereafter, the cells were washed once with growth media and twice with DPBS. Labeled cells were imaged as described above. The coumarin filter set was 405/20 excitation, 425 dichroic, 460/40 emission.

References

1. Agard,N.J., Baskin,J.M., Prescher,J.A., Lo,A. & Bertozzi,C.R. A comparative study of bioorthogonal reactions with azides. *ACS Chem. Biol.* **1**, 644-648 (2006).

2. Agard,N.J., Prescher,J.A. & Bertozzi,C.R. A strain-promoted [3 + 2] azide-alkyne cycloaddition for covalent modification of biomolecules in living systems. *J. Am. Chem. Soc.* **126**, 15046-15047 (2004).

3. Griffin,R.J. The medicinal chemistry of the azido group. *Prog. Med. Chem.* **31**, 121-232 (1994).

4. Adams,S.R. *et al.* New biarsenical ligands and tetracysteine motifs for protein labeling in vitro and in vivo: synthesis and biological applications. *J. Am. Chem. Soc.* **124**, 6063-6076 (2002).

5. Willnow,T.E. The low-density lipoprotein receptor gene family: multiple roles in lipid metabolism. *J. Mol. Med.* **77**, 306-315 (1999).

6. Nieland,T.J., Ehrlich,M., Krieger,M. & Kirchhausen,T. Endocytosis is not required for the selective lipid uptake mediated by murine SR-BI. *Biochim. Biophys. Acta* **1734**, 44-51 (2005).

7. Baskin,J.M. *et al.* Copper-free click chemistry for dynamic in vivo imaging. *Proc. Natl. Acad. Sci. U. S. A* **104**, 16793-16797 (2007).

8. Chen,I., Howarth,M., Lin,W. & Ting,A.Y. Site-specific labeling of cell surface proteins with biophysical probes using biotin ligase. *Nat. Methods* **2**, 99-104 (2005).

9. Lin,C.W. & Ting,A.Y. Transglutaminase-catalyzed site-specific conjugation of small-molecule probes to proteins in vitro and on the surface of living cells. *J. Am. Chem. Soc.* **128**, 4542-4543 (2006).

10. Nauman,D.A. & Bertozzi,C.R. Kinetic parameters for small-molecule drug delivery by covalent cell surface targeting. *Biochim. Biophys. Acta* **1568**, 147-154 (2001).

11. Howarth,M., Takao,K., Hayashi,Y. & Ting,A.Y. Targeting quantum dots to surface proteins in living cells with biotin ligase. *Proc. Natl. Acad. Sci. U. S. A* **102**, 7583-7588 (2005).

12. Howarth,M. *et al.* A monovalent streptavidin with a single femtomolar biotin binding site. *Nature Methods* **3**, 267-273 (2006).

13. Howarth,M. *et al.* Monovalent, reduced-size quantum dots for imaging receptors on living cells. *Nat. Methods* **5**, 397-399 (2008).

14. Howarth,M. & Ting,A.Y. Imaging proteins in live mammalian cells with biotin ligase and monovalent streptavidin. *Nat. Protoc.* **3**, 534-545 (2008).

15. Reche,P. & Perham,R.N. Structure and selectivity in post-translational modification: attaching the biotinyl-lysine and lipoyl-lysine swinging arms in multifunctional enzymes. *EMBO J.* **18**, 2673-2682 (1999).

16. Pasquale,E.B. Eph receptor signalling casts a wide net on cell behaviour. *Nat. Rev. Mol. Cell Biol.* **6**, 462-475 (2005).

17. Singh,A.B. & Harris,R.C. Autocrine, paracrine and juxtacrine signaling by EGFR ligands. *Cell Signal.* **17**, 1183-1193 (2005).

18. Tuli,S.S. *et al.* Immunohistochemical localization of EGF, TGF-alpha, TGF-beta, and their receptors in rat corneas during healing of excimer laser ablation. *Curr. Eye Res.* **31**, 709-719 (2006).

19. Flanagan,J.G. & Vanderhaeghen,P. The ephrins and Eph receptors in neural development. *Annu. Rev. Neurosci.* **21**, 309-345 (1998).

20. Wimmer-Kleikamp,S.H. & Lackmann,M. Eph-modulated cell morphology, adhesion and motility in carcinogenesis. *IUBMB. Life* **57**, 421-431 (2005).

21. Kim,D.J. *et al.* Crystal structure of lipoate-protein ligase A bound with the activated intermediate: insights into interaction with lipoyl domains. *J. Biol. Chem.* **280**, 38081-38089 (2005).

22. McManus,E., Luisi,B.F. & Perham,R.N. Structure of a putative lipoate protein ligase from Thermoplasma acidophilum and the mechanism of target selection for post-translational modification. *J. Mol. Biol.* **356**, 625-637 (2006).

23. Ali,S.T. & Guest,J.R. Isolation and characterization of lipoylated and unlipoylated domains of the E2p subunit of the pyruvate dehydrogenase complex of Escherichia coli. *Biochem. J.* **271**, 139-145 (1990).

24. Gama,L. & Breitwieser,G.E. Generation of epitope-tagged proteins by inverse polymerase chain reaction mutagenesis. *Methods Mol. Biol.* **182**, 77-83 (2002).

Chapter 8 : Intracellular protein labeling with lipoic acid ligase

The work discussed in this chapter is unpublished. The cyclooctyne-fluorophore conjugates and coumarin probes were synthesized by Dr. Hemanta Baruah.

Introduction

In Chapter 7 we demonstrated that LplA-catalyzed labeling offers an excellent platform for sensitive and specific cell-surface protein labeling. Although LplA offered certain advantages over previously developed cell-surface labeling methods, such as biotin ligase and transglutaminase, what excited us the most about this new methodology was its potential to work inside cells. We described in the introductory Chapter 5 that none of the currently available methods for genetic targeting of chemical probes has been able to accomplish intracellular labeling of proteins fused to just a small peptide tag. For example, SNAP-tag[1] and HaloTag[2] allow labeling of intracellular proteins but they both impose a large tag on the target protein (182 and 296 amino acid residues, respectively). In contrast, biotin ligase[3] and phosphopantetheine transferase[4] recognize 15- and 11-mer peptide tags, but they are both restricted to the cell-surface. As we have shown, *E. coli* lipoic acid ligase (LplA) recognizes a 22-amino acid peptide recognition sequence (the LAP peptide) and can ligate small-molecule probes such as a coumarin fluorophore or an alkyl azide functional group handle. On one hand, we expect the coumarin derivative to be membrane permeable because it is small and uncharged and because other derivatives of coumarin have been used in the intracellular milieu before[5,6]. On the other hand, azides can chemoselectively react with alkynes via the [3+2] cycloaddition. Both azides and alkynes are abiotic and possess extremely selective reactivity toward each other; thus, inside the cells, they will only react with each other. Although there had been no previous reports of intracellular strain-promoted [3+2] cycloaddition, we expected the alkyl azide to be membrane pemeant, and we could design the cyclooctyne conjugate to be so, as well.

Figure 8-1 depicts the general scheme for two-step intracellular labeling. In contrast to cell-surface labeling, where only the LAP-tagged protein needed to be transfected, for intracellular labeling, both the LAP-fused protein and the LplA enzyme are co-transfected and expressed in the cytosol (or other subcellular compartment) of mammalian cells. Azide **7** is then added in the cell growth media, and we expect it to passively diffuse through the plasma membrane, and once inside the cell, be ligated by LplA, which uses endogenous ATP. Thereafter, excess azide **7** can be removed from the cells by extensive washes before adding the membrane-permeant

cyclooctyne-fluorophore conjugate. Intracellular [3+2] cycloaddition results in fluorescent tagging of only the LAP-tagged proteins.

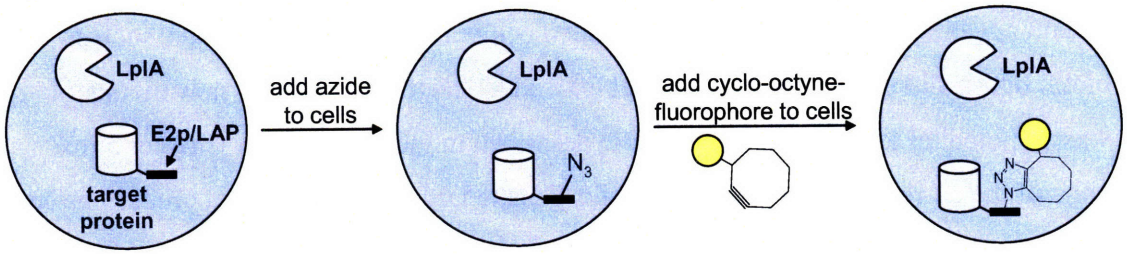


Figure 8-1: LplA-mediated two-step labeling of intracellular proteins. Cells are first co-transfected with plasmids encoding a LAP- or E2p-fused protein and the lipoic acid ligase (LplA) enzyme. The first step of the labeling is the enzymatic ligation of a membrane-permeant alkyl azide probe (azide **7**) by the intracellular LplA, which uses endogenous ATP. The second step of the labeling is the chemoselective reaction of azide-derivatized proteins with a membrane-permeant cyclooctyne-fluorophore conjugate to form a stable cycloadduct between the target protein and the fluorophore.

Results

LplA catalyzes the ligation of azide 7 to E2p inside mammalian cells

We first wished to demonstrate the first step of LplA-catalyzed intracellular ligation of azide **7**. In order to maximize our labeling signal, initial proof of concept experiments were performed on the hybrid lipoyl domain derived from the second subunit of the *E. coli* pyruvate dehydrogenase complex (the E2p domain) because the kinetics of the LplA-catalyzed ligation of azide **7** are faster than those of the engineered LplA acceptor peptide (LAP). We created plasmids for intracellular expression of E2p and LplA and we expressed them in human cervical carcinoma (HeLa) cells. Cells were grown in normal cell growth media supplemented with 750 μM azide **7** for ~ 48 hours after transfection. We had previously determined, using a mitochondrial respiration assay, that 48-hour exposure to azide **7** had no effect in cell health at concentrations equal or less than 750 μM (data not shown). HeLa cells were lysed and the ligated azide **7** was detected using a mobility shift assay. This assay had been previously developed by Ali

and Green and it is based on the fact that modified E2p moves faster than the apo protein in a native gel owing to the loss of the positively charged lysine[7]. E2p was detected by blotting its N-terminal hemagglutinin (HA) epitope with an anti-HA antibody. The results shown in Figure 8-2 demonstrated that intracellular ligation of azide **7** by LplA was possible. Lane 1 shows the mobility of the apo protein, which is different from the mobility of the E2p protein after co-expression with LplA and in-cell treatment with azide **7** (lane 6). Lane 2 is a positive control, where E2p is modified *in vitro* with azide **7** by purified LplA, and it shows the mobility shift of the E2p domain upon modification of its lysine site of attachment. Lanes 3 and 4 demonstrate that the E2p domain is not modified by the mammalian lipoic acid ligase (LPL) neither in cells nor *in vitro*. Lane 6 shows that E2p is not modified by LplA unless azide **7** is added to the medium, that is, LplA does not ligate endogenous lipoic acid to the E2p substrate. This is probably a reflection of the much higher concentration of azide **7** present in the cells over that of the endogenous lipoic acid.

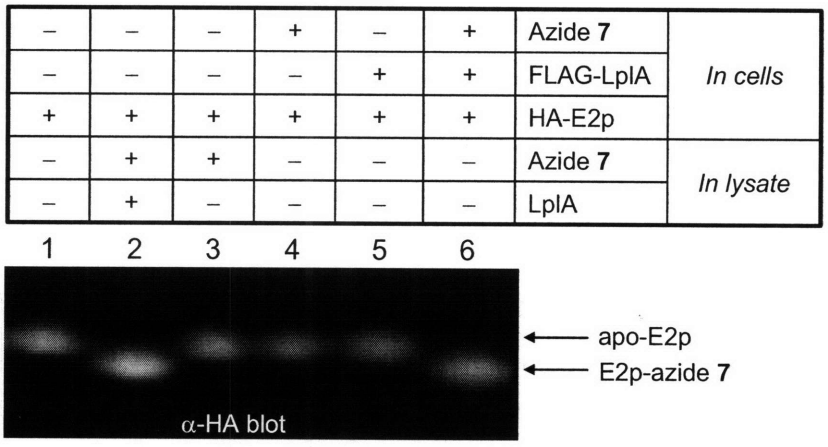


Figure 8-2: Intracellular ligation of azide 7 detected by a gel shift assay. HEK cells expressing intracellular LplA (pcDNA3-FLAG-LplA) and E2p (pcDNA3-HA-E2p) were grown in the presence of 750 μM azide **7** for ~ 48 hours before lysis. Cell extracts were run in a native 20% PAGE and the E2p domain was detected by anti-HA blotting. Lane 1 shows the mobility of unmodified E2p domain. Lane 6 demonstrates intracellular ligation of azide **7** by LplA. Negative controls are shown with LplA or azide **7** omitted (lanes 4 and 5, respectively). Additional controls are shown with azide **7** omitted from the cell labeling but added later to the lysates with or without addition of purified LplA (lanes 2 and 3).

Overall, this experiment confirmed three essential requirements for LplA-mediated intracellular labeling: that the azide **7** is membrane permeant, that the E2p substrate is not pre-modified with lipoic acid by neither LplA nor the mammalian LPL,

and that the transfected LplA is active and can ligate azide **7** to E2p in the cytosol of mammalian cells. Importantly, we found that, under the above conditions, the intracellular ligation of azide **7** ligation to E2p displayed a 100% yield. However, two other important requirements were not tested by the gel-shift assay. First, the gel shift assay did not provide information on whether other proteins in the cells were also being modified by LplA. In Chapters 6 and 7, we demonstrated that LplA is extremely specific toward its protein or peptide substrates both in the context of a cell lysate (Figure 6-12) and at the surface of mammalian cells (Figures 7-4 and 7-5). Here, we wished to confirm that LplA retained such high specificity when expressed in the cytosol of mammalian cells. Second, because the mobility shift is caused by the loss of the positively charged lysine, it was unclear whether it would distinguish between the ligation of an intact azide or an azide that had been reduced to a primary amine. Azide reduction to amines had been previously reported to occur in the cytosol of mammalian cells and it was therefore a major concern for us.

We thus tested the sequence specificity of LplA, and the retention of azide reactivity of the ligated azide **7** with a western blot assay similar to the one already described in Chapter 6 (Figure 6-12). Briefly, human embryonic kidney (HEK) cells expressing intracellular E2p and LplA were grown in normal cell growth media supplemented with 750 μM azide **7** for ~ 48 hours after transfection. Thereafter, cells were lysed and azide-tagged proteins were derivatized *in vitro* via Cu(I)-mediated [3+2] cycloaddition using a biotin-alkyne probe[8]. Ligated biotin-alkyne was detected by streptavidin blotting. Figure 8-3 shows the streptavidin stained western blot. The results suggested that, similar to Figure 6-12, in the presence of thousands of mammalian proteins in the cell lysate, only E2p-CFP was labeled by LplA (lane 2). We performed negative controls with azide **7** omitted (lane 1), the lysine site of E2p modification mutated to alanine (E2p(ala)-CFP, lane 4), or with wild-type LplA replaced by a catalytically inactive mutant of LplA (LplA(arg), lane 4), and saw no labeling in all cases. These results thus suggested that LplA retained its specificity when expressed inside cells and that the azide **7** probe was not, at least in part, being reduced to an amine.

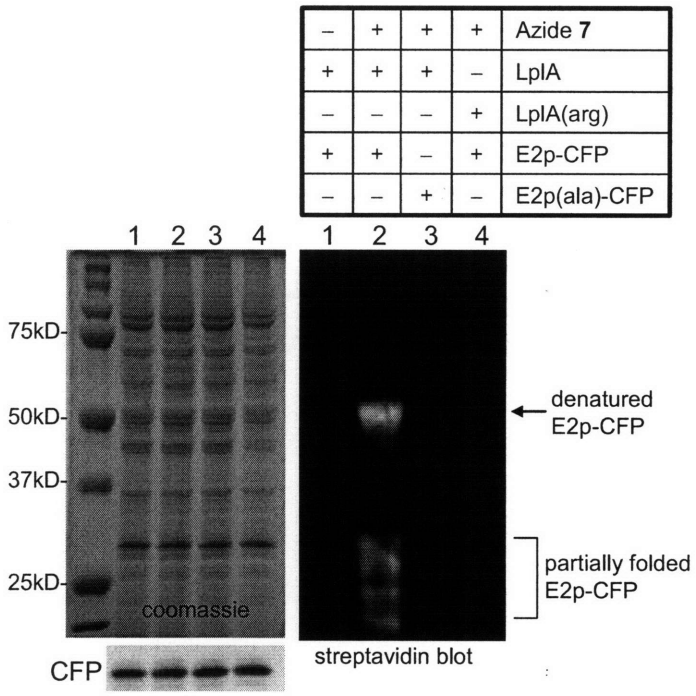
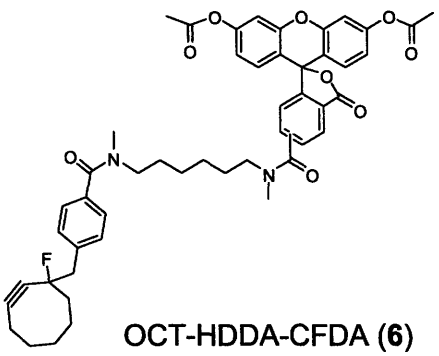


Figure 8-3: Intracellular ligation of azide 7 detected by secondary reaction with biotin-alkyne. HEK cells expressing intracellular LplA (pcDNA3-FLAG-LplA) and E2p (pcDNA3-E2p-CFP) were grown in the presence of 750 μM azide **7** for ~ 48 hours before lysis. The azide was derivatized *in vitro* with biotin-alkyne in the presence of copper catalyst via the [3+2] cycloaddition and the ligated biotin was detected by streptavidin blotting. Controls are shown with azide **7** omitted (lane 1), with E2p-CFP replaced by its alanine point mutant (lane 3), or with LplA replaced by its catalytically inactive Lys133Arg mutant (lane 4). Coomassie staining demonstrates equal loading in all lanes. Fluorescence visualization of CFP demonstrates equal expression levels of the E2p fusion in all lanes.

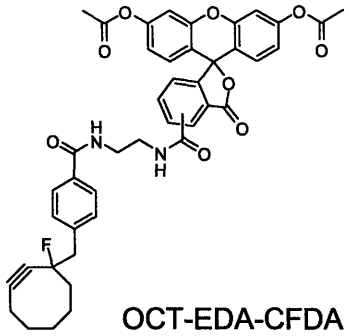
Synthesis and testing of membrane-permeant cyclooctyne conjugates to detect azide 7 ligation inside live cells

Next, we wished to detect the ligation of azide **7** directly in living cells via the strained-promoted [3+2] cycloaddition. We synthesized and tested a panel of cyclooctyne conjugates to several organic fluorophores. The structures of some of these probes are shown in Figure 8-4. All the OCT-fluorophore probes were synthesized by Dr. Hemanta Baruah following a scheme analogous to the one described for the OCT-PEG-fluorophore conjugates in Chapter 7 (Figure 7-2). In contrast to our cell-surface OCT-PEG-fluorophore conjugates, the intracellular labeling probes needed to be membrane-permeant, such that they could passively diffuse inside the cells, where they react with azide-derivatized proteins. Excess unreacted probe could then either diffuse out or be

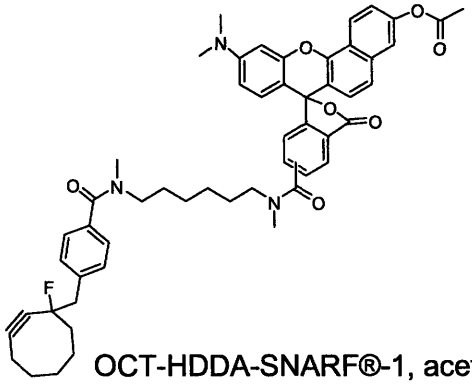
pumped-out by non-specific endogenous anion transporters, such as the human organic anion transporters (hOATs)[9] or the ATP-binding cassette (ABC) transporters[10].



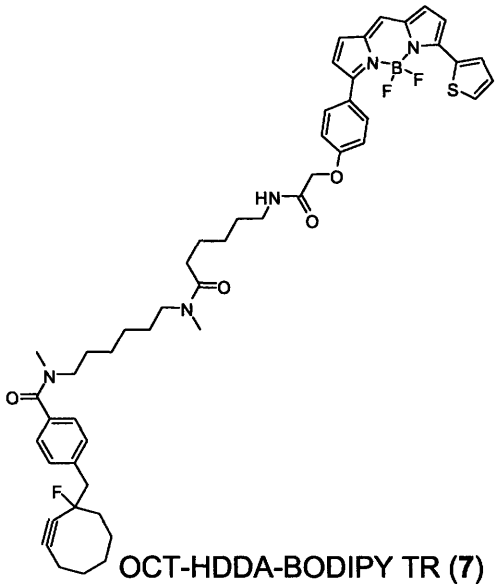
OCT-HDDA-CFDA (**6**)



OCT-EDA-CFDA (**10**)



OCT-HDDA-SNARF®-1, acetate (**9**)



OCT-HDDA-BODIPY TR (**7**)

Figure 8-4: Structures of OCT-fluorophores conjugates tested for intracellular labeling applications.

We chose carboxyfluorescein diacetate (CFDA)[1,11-14], SNARF®-1-acetate[15], and the green and red BODIPY dyes (BODIPY-Fl and BODIPY-TR)[14] because they had all been previously used in intracellular labeling applications. As depicted in Figure 8-4, the hydroxyl groups in carboxyfluorescein and SNARF®-1 were acetate-protected to render the molecules neutral, thus capable of crossing the cell membrane. Once in the cytosol, the acetate groups are readily hydrolyzed, probably by non-specific endogenous esterases[6]. We also compared two different linkers: a short ethylenediamine (EDA) and a longer more hydrophobic *N,N'*-dimethyl-1,6-diaminohexane (HDDA).

We tested the cyclooctyne derivatives for compliance with our two main requirements: fast kinetics of loading into the cells and coming out of the cells, and minimal background staining. HeLa cells were incubated for 5–60 minutes with different concentrations of each probe at different temperatures. After loading, the probes were allowed to be released from the cells for 15–120 minutes. We imaged the cells before and after the washes. Table 8-1 summarizes the results.

Table 8-1: Comparison of several OCT-fluorophore conjugates for intracellular labeling applications. CFDA: 5/6-carboxyfluorescein diacetate. HDDA: *N,N'*-dimethyl-1,6-diamino hexane. EDA: ethylenediamine. BODIPY-FL: green BODIPY. BODIPY-TR: red BODIPY.

Fluorophore	Linker	Endocytosis	Optimal Concentration (μM)	Time IN (minutes)	Time OUT (minutes)	Staining pattern
CFDA	HDDA	No at T < 30 °C	1–5	5–15	30–45	Diffuse cytosolic
CFDA	EDA	No at T < 30 °C	5	30	60–90	Diffuse cytosolic
BODIPY-FL	HDDA	No	0.1	5–15	No	Punctate (mitochondrial)
BODIPY-FL	EDA	No	0.1	5–15	No	Punctate (mitochondrial)
BODIPY-TR	HDDA	Yes at T > 16 °C	0.1	5	No	Punctate (mitochondrial)
SNARF®-1	HDDA	No at T < 30 °C	5	30	No	Mitochondrial and perinuclear

We observed very different loading kinetics and staining patterns for all the probes tested. BODIPY dyes, for example, diffused inside the cells most efficiently, requiring only 5 minutes of loading at a very low dye concentration (100 nM). They did not, however, wash out of the cells even after 2 hours. Their staining pattern was highly punctated and excluded from the nucleus. Co-localization with the mitrochondrial stain MitoTracker Red suggested that the OCT-BODIPY probes were non-specifically bound to the mitochondrial membrane. A similar staining pattern was observed for OCT-SNARF®-1, except that its non-specific staining pattern was more general, co-localizing also with the endoplasmic reticulum and the golgi apparatus. Additionally, OCT-SNARF®-1 required higher dye concentrations and longer loading times than BODIPY dyes. In contrast, OCT-HDDA-CFDA conjugates were able to both diffuse in and wash

out of cells, with OCT-HDDA-CFDA displaying better kinetics than OCT-EDA-CFDA. Additionally, because endocytosis, but not passive diffusion, is blocked at 4 °C, we used the temperature dependence of the loading as well as the staining pattern (i.e., punctate if endocytosed) to determine whether a probe was being uptaken by endocytosis or not. Importantly, the OCT-HDDA-CFDA probes did not exhibit any endocytosis at temperatures lower than 30 °C. It must be noted that addition of 0.1% pluronic, a nonionic surfactant, aided the loading of all the probes tested.

We thus chose to use the OCT-HDDA-CFDA conjugate to test LplA-mediated intracellular protein labeling in live cells. As schematically shown in Figure 8-1, HeLa or HEK cells expressing intracellular E2p-CFP and LplA were first grown for 24 hours in the presence of 500 μM azide **7**. After allowing the excess azide **7** to passively diffuse out of the cells (~ 1 hour), azide-modified proteins were derivatized by incubation for 1 hour with 5 μM OCT-HDDA-CFDA. Initial tests in HeLa cells resulted in non-detectable signal. Increasing the labeling time with OCT-HDDA-CFDA alone did not help, because the acetate protecting groups of CFDA were cleaved in the cell media after ~ 1 hour, thus rendering the molecule membrane impermeant. Addition of 2.5 mM probenecid, an inhibitor of the anion transporters that pump carboxyfluorescein out of the cells, did however increase the labeling signal to a level that was clearly detectable by imaging (data not shown). By using probenecid, OCT-HDDA-CFDA could be retained inside cells for 2–4 hours, thus allowing it enough time to react. In HEK cells, the labeling signal was a lot higher, even in the absence of probenecid, presumably because of the higher transfection and protein expression levels. Figure 8-5 illustrates a typical result of live-cell detection of azide **7** ligation in HEK using OCT-HDDA-CFDA.

The images were striking. Although we observed a signal (fluorescein panel) that was dependent on azide **7** treatment and on LplA enzymatic activity, the signal intensity was comparable on cells expressing E2p-CFP or its alanine mutant (E2p(ala)-CFP). Additional experiments confirmed that the labeling persisted even in the absence of E2p (data not shown). These results, which suggested that LplA was labeling other proteins inside the cells, were in stark contrast with the specificity test shown in Figure 8-3. The fact that the background staining was dependent on LplA activity ruled out the possibility that the non-specific labeling was generated by non-specific reactivity of the

OCT-HDDA-CFDA probe. The western blot results also made it unlikely that the background labeling was being originated by LplA self-labeling or labeling of the four mammalian lipoate-acceptor proteins, because such a high background signal (as seen in Figure 8-5), if concentrated on a single or four other proteins, should have been clearly detectable by western blot.

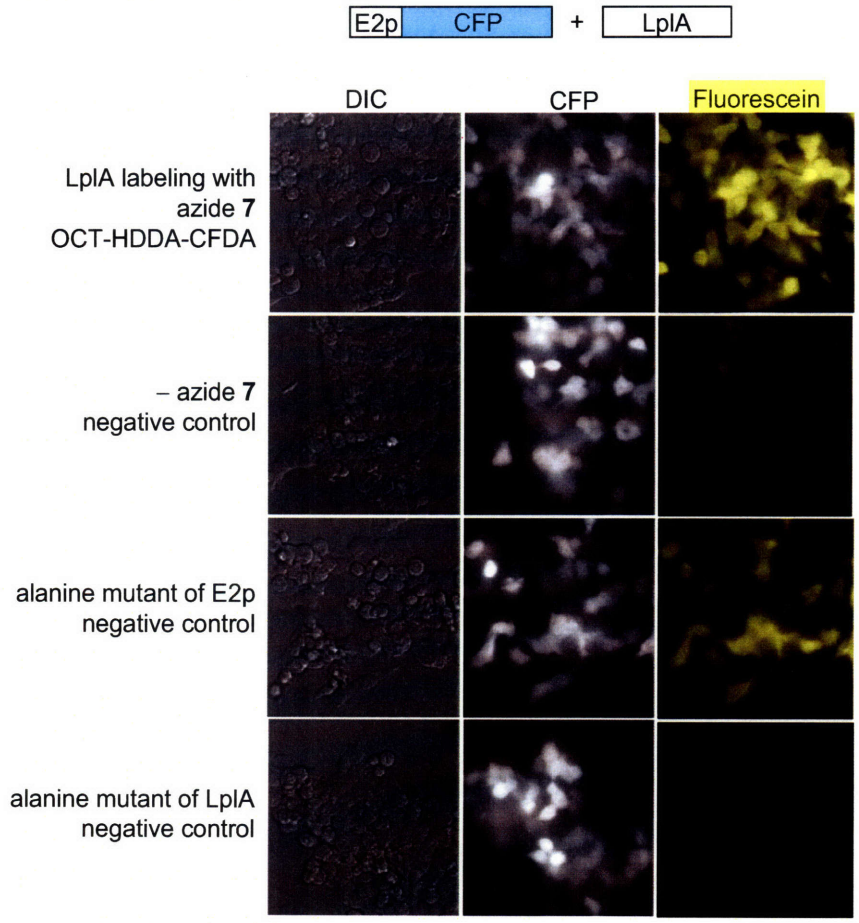


Figure 8-5: Background staining during the live-cell detection of intracellular ligation of azide 7 using OCT-HDDA-CFDA. An E2p-CFP fusion and LplA were expressed in the cytosol of HEK cells. Intracellular E2p was first labeled with azide **7** by the expressed LplA, and the introduced azide was then derivatized with a cyclooctyne probe conjugated to carboxyfluorescein diacetate (OCT-HDDA-CFDA). Live cell images of the introduced fluorescein are shown to the right of the CFP and DIC images, which highlight the transfected cells. Negative controls with azide **7** omitted from the cell media, with E2p replaced by its alanine point mutant, or with LplA replaced by a catalytically inactive mutant (Lys133Ala) are shown.

We reasoned that the discrepancy between the western blot and imaging results was caused by the fact that, when western blot was used as a readout, the background

labeling of many other endogenous proteins was distributed among the different molecular weights. Therefore, labeling of E2p dominated the image simply because it was concentrated at a single molecular weight. In contrast, when imaging was used as a readout, all the proteins were indistinguishably expressed in the cytosol, and thus the E2p signal was not significantly higher than the sum of the signal of other labeled endogenous proteins. We will later describe how this hypothesis was confirmed when we performed more careful comparisons using western blot detection (see Figure 8-8 and Figure 8-10).

Attempts to reduce background staining by lowering the expression levels of intracellular LplA (i.e., by decreasing the amount of DNA used during transfection) and by decreasing the azide **7** concentration and labeling time were all unsuccessful. We did however found, during our troubleshooting, that the enzyme was not expressing properly inside cells. When we performed fixed-cell immunostaining of HeLa cells expressing LplA, we observed that the enzyme was highly aggregated. In fact, LplA staining was concentrated on strikes (or sometimes in large vesicles) in the perinuclear region (Figure 8-6). We thus hypothesized that enzyme overexpression and aggregation was the main cause of LplA's loss of ligation specificity.

In the next sections, we describe our iterative process to improve LplA folding and expression inside cells, which ultimately resulted in the recovery of its labeling specificity.

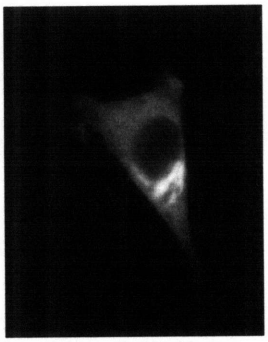


Figure 8-6: Fixed-cell immunodetection of LplA expressed inside HeLa cells. HeLa cells were co-transfected with E2p-CFP and LplA. 48 hours after transfection, cells were fixed and permeabilized with formaldehyde/methanol. Anti-FLAG staining detects expression of FLAG-LplA.

Improving expression and folding of LplA inside mammalian cells

We attempted to improve enzyme expression by fusing LplA to the fluorescence protein mCherry. Fusion of hard-to-express or unstable proteins to certain fluorescence proteins had been previously reported to improve protein expression and solubility[16]. We chose the mCherry protein because its excitation and emission spectra are well-separated from those of CFP and fluorescein. Although fusion of LplA to mCherry resulted in a marked increase in the expression levels of intracellular LplA (as assessed by anti-FLAG blotting) and partially reduced aggregation, the background staining upon azide **7** and OCT-HDDA-CFDA labeling remained almost unchanged (data not shown). It is worth mentioning that expression of the mCherry-LplA construct in HeLa (but not HEK) cells, where enzyme expression levels were lowest, sometimes but not reproducibly, resulted in specific OCT-HDDA-CFDA labeling when short azide **7** labeling times were employed. The signal obtained was however very low (signal/background ratio < 1.5) and thus hardly reproducible.

We also created a series of LplA mutants, where we mutated each LplA surface cysteine to serine. The rationale behind this approach was that perhaps aggregation was being caused in the oxidizing environment of the endoplasmic reticulum because of enzyme oligomerization via disulfide-bonding. Cysteines 22, 55, 200, 300, and 311 were mutated to serines, and the intracellular expression levels and aggregation patterns of the single-point mutants were visualized by anti-FLAG immunostaining. Cys200Ser and Cys330Ser mutants behaved analogously to the wild-type enzyme in terms of expression levels and background staining with OCT-HDDA-CFDA, whereas Cys22Ser did not express at all. In contrast, both Cys55Ser and Cys311Ser seemed to retain high expression levels while significantly reducing enzyme aggregation, although this reduction was not accompanied by a decrease in background OCT-HDDA-CFDA staining.

In parallel, we also created mutants of Trp37. As described in Chapter 7, Trp37 lays in the crystal structure of *Thermoplasma acidophilum* (Ta) LplA in close proximity to the lipoic acid dithiolane ring. Dr. Hemanta Baruah in our lab had found that mutation of this gatekeeper residue to either valine or isoleucine resulted in an LplA mutant capable of efficient ligation of a coumarin derivative. Additionally, mutation of the same

Trp37 to serine enabled incorporation of an aryl azide photoaffinity probe. We tested the intracellular activity of the W37I, W37V, and W37S mutants fused to mCherry, and we were surprised to observe that the three mutants behaved dramatically different. Whereas W37S expression levels were very similar to those of the wild-type enzyme, W37V expression was highly aggregated in the form of large puncta (similar to the original FLAG-LplA), and W37I expression levels were too low to be detected by imaging. It must be noted that the three mutants behaved similarly well when overexpressed in *E. coli*.

Taken together, the results of the Cys-to-Ser mutants and the W37 mutant series suggested that the intracellularly expressed LplA enzyme was so unstable that a single point mutant, even if buried in the enzyme active site, could cause unfolding and aggregation or even completely abolish enzyme expression. This problem, although somewhat alleviated, was not completely solved by fusing the enzyme to mCherry.

We thus decided to express LplA from a gene sequence optimized for expression in human cell lines. We purchased the synthetic codon-optimized gene and we cloned it as a fusion to mCherry for intracellular expression. To differentiate it from the enzyme expressed from the original *E. coli* gene, we called the enzyme expressed from the codon-optimized gene, hLplA, which stands for humanized LplA. We transfected HEK cells with the new mCherry-hLplA construct together with E2p-CFP and we performed the azide **7** labeling followed by detection with OCT-HDDA-CFDA as described above. The results were, once again, striking, as they suggested that expression of LplA from the codon-optimized had finally recovered the loss of labeling specificity (Figure 8-7). We observed what seemed specific labeling of transfected (CFP- and mCherry-positive) cells, without labeling of neighboring untransfected cells in the same field of view. More importantly, the negative control where E2p-CFP was replaced by its alanine point mutant (alanine at the lysine modification site within E2p) showed no labeling, which suggested that the ligation of azide **7** was site-specific. Interestingly, enzyme expression from the codon optimized gene resulted in much higher intracellular concentrations, as assessed by the mCherry signal. Aggregation was also almost eliminated, except for a few cells that still displayed few brighter rounded dots (indicated by white arrows in Figure 8-7). We later found, however, that the labeling results were a mere artifact

introduced by the fact that the enzymatic activity of hLplA was inhibited by the alanine mutant of E2p. hLplA was however not inhibited by the alanine mutant of the LAP peptide, and labeling performed in cells that expressed either LAP(ala) or no protein or peptide substrate at all resulted, once again, in high background staining. To date, we have not found a plausible explanation for why hLpla but not eLplA (expressed from the *E. coli* gene) displays enzyme inhibition by E2p(ala).

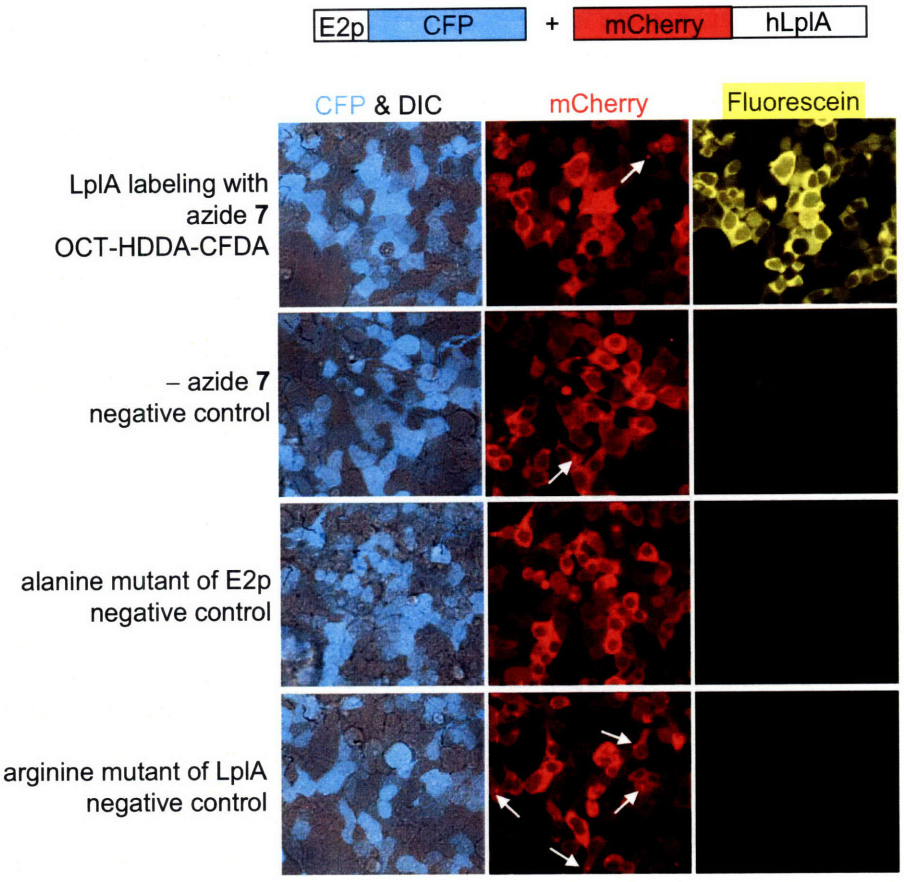


Figure 8-7: Live-cell detection of intracellular ligation of azide 7 by a codon-optimized LplA enzyme using OCT-HDDA-CFDA. E2p-CFP and mCherry-hLplA fusions were expressed in the cytosol of HEK cells. Intracellular E2p was first labeled with azide **7** by the expressed hLplA, and the introduced azide was then labeled with a cyclooctyne probe conjugated to carboxyfluorescein diacetate (OCT-HDDA-CFDA). Live cell images of the introduced fluorescein are shown to the right of the merged CFP and DIC, and the mCherry images, which highlight the transfected cells. Negative controls with azide **7** omitted from the cell media, with E2p-CFP replaced by its alanine point mutant, or with LplA replaced by a catalytically inactive mutant (Lys133Arg) are shown. White arrows point to spots cells were the enzyme is partially aggregated in the form of brighter dots.

The confusion introduced by the previous imaging results motivated us to embark in a systematic study of the ligation specificity of all the LplA constructs, with the ultimate goal of elucidating the source of background staining. We first tested the specificity of the intracellular ligation of the natural substrate lipoic acid, both by western blot and by imaging. We then compared the results with those obtained for the ligation of azide **7**. We also created an additional construct, by fusing LplA (both, eLplA and hLplA) to the FKBP-rapamycin binding domain of the mTor protein (FRB), with the intention of promoting enzyme turnover. If, as we hypothesized, the loss of specificity tightly correlated with the enzyme expression levels, we reasoned that introducing protein turnover could perhaps introduce a second layer of regulation that resulted in more physiological enzyme concentrations. Table 8-2, at the end of this section, summarizes the results of the comparative study. In the following experiments, we refer to the LplA enzyme expressed from the *E. coli* gene as eLplA to differentiate it from hLplA. It must be noted that eLplA is the same enzyme that we have been thus far calling simply LplA.

First, to compare the specificity and yield of intracellular lipoic acid ligation, we co-transfected HEK cells with either E2p-CFP or LAP-CFP and each one of the LplA constructs. Cells were grown in the presence of 500 μM lipoic for 24 hours, and ligated lipoic acid was detected by anti-lipoic western blotting. Figure 8-8 shows the results. We observed that, whereas all constructs were able to efficiently lipoylate E2p, expression of LplA from both hLplA constructs resulted in high background labeling of many other proteins present in the lysate (seen as a smear in Figure 8-8). It was also interesting to see that hLplA expression achieved much larger yields of LAP labeling than eLplA. The observed LAP-CFP labeling was mostly site-specific, as it was not seen in the negative control where cells were transfected with the alanine mutant of LAP. We observed no difference in the labeling specificity or yield between the corresponding mCherry and FRB fusions.

We were very surprised to find such a stark difference in labeling specificity between the eLplA and hLplA constructs. We rationalized this difference based on the increased expression levels of the hLplA enzyme rather than on an intrinsic property. This hypothesis was later confirmed by Dr. Shah, who cloned the mCherry-hLplA construct downstream of a tetracycline operator for inducible protein expression. As

explained in Chapter 4, the tetracycline regulation in the T-REx™ System is based on the binding of tetracycline to the Tet repressor, which releases the repressor and results in derepression of the promoter controlling expression of the gene of interest[17]. Thus, increasing concentrations of tetracycline result in increased protein expression. Dr. Shah demonstrated that mCherry-hLplA can in fact achieve highly specific ligation of lipoic acid when it is expressed at very low concentrations (~ 0–0.03 ug/mL tetracycline; ~ 10-fold lower expression than that obtained from the non-tetracycline regulated plasmid).

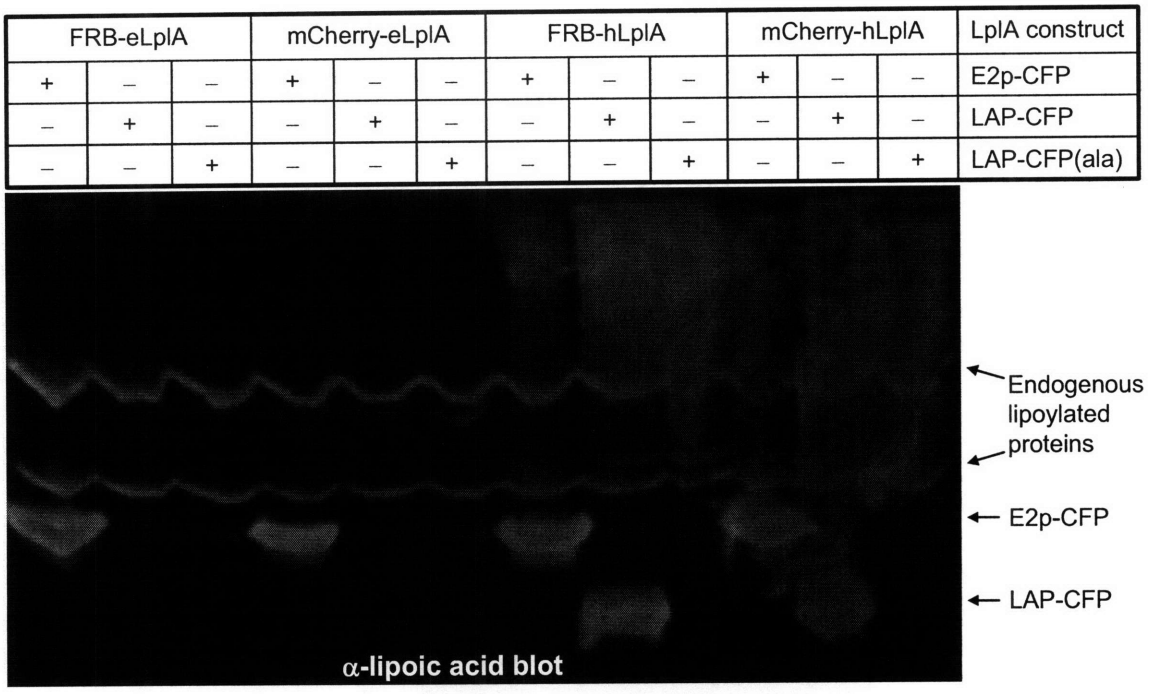


Figure 8-8: Western blot detection of LplA-mediated intracellular ligation of lipoic acid. HEK cells were transfected with constructs for expression of intracellular E2p-CFP or LAP-CFP together with one of the four LplA fusion constructs (FRB-eLplA, FRB-hLplA, mCherry-eLplA, or mCherry-hLplA). All cells were then grown in the presence of 500 μM lipoic acid for ~ 24 hours before lysis. Ligated lipoic acid was detected in lysates by anti-lipoic blotting. A negative control with E2p and LAP replaced by the alanine mutant of LAP (LAP(ala)) is also shown. We used LAP(ala) as the negative control instead of alanine mutant of E2p because of the demonstrated enzyme inhibition of hLplA by E2p(ala). eLplA stands for the enzyme expressed from the original *E. coli* lpla gene. hLplA stands for the enzyme expressed from a gene sequence optimized for expression in human cell lines.

To further confirm the specificity of mCherry-eLplA (and FRB-eLplA) we repeated the same assay for intracellular lipoic acid ligation with the inclusion of additional controls. The results seen in Figure 8-9 confirmed that the intracellular ligation of lipoic acid by mCherry-LplA is site-specific (lane 2) and depends on the enzymatic

activity of the transfected LplA (lane 3). This means that the mammalian LPL is not capable of ligating azide **7** to E2p. Furthermore, the lipoylation levels of the endogenous lipoate-acceptor proteins did not change upon transfection with LplA (compare lanes 2 and 4), which confirmed that cytosolic LplA did not label endogenous lipoate-acceptor proteins. This property may be a reflection of LplA not being expressed in the mitochondria, where the endogenous lipoylated proteins are located. It remains to be seen if the specificity holds on if LplA is targeted for mitochondrial expression. Additionally, treatment of untransfected cells with lipoic acid did not affect the basal levels of lipoylation of the endogenous lipoate-acceptor proteins (compare lanes 4 ad 5), suggesting that the labeling procedure should not be toxic to cells. This is probably true because of the biological function of the mammalian LPL proteins substrates, which requires that the proteins be 100% lipoylated at all times, and because of the slow turnover characteristic of these proteins.

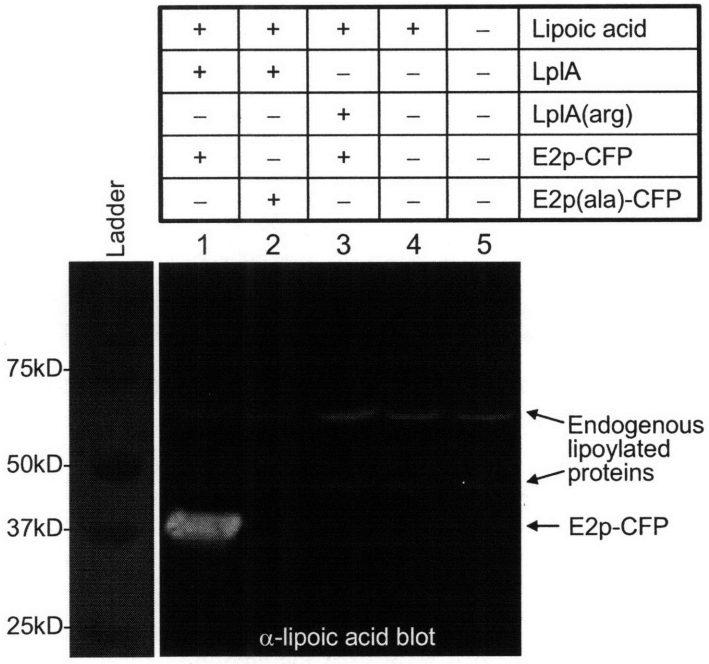


Figure 8-9: Intracellular ligation of lipoic acid by mCherry-LplA is specific and orthogonal in mammalian cells. HEK cells expressing intracellular LplA (pcDNA3-mCherry-LplA) and E2p-CFP (pcDNA3-E2p-CFP) were grown in the presence of 500 μM lipoic acid for ~ 24 hours before lysis. Ligated lipoic acid was detected by anti-lipoic acid blotting. Controls are shown with E2p-CFP replaced by its alanine point mutant (lane 2), or with LplA replaced by its catalytically inactive Lys133Arg mutant (lane 3). Lysates from untransfected cells, treated with lipoic acid (lane 4) or left untreated (lane 5) are also shown.

In summary, this assay unequivocally demonstrated that, if expressed at physiological concentrations, the intrinsic specificity of LplA is good enough for intracellular ligation of, at least, its natural substrate lipoic acid. Furthermore, the labeling is orthogonal as it is not affected by the endogenous mammalian lipoylation activity.

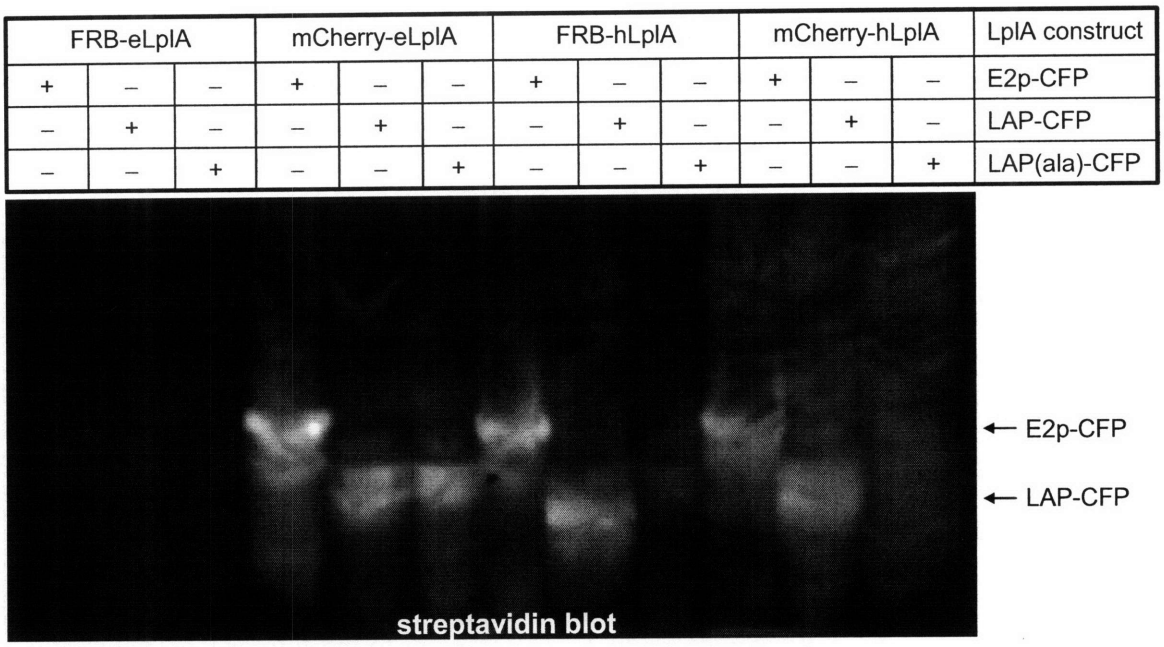


Figure 8-10: Western blot detection of LplA-mediated intracellular ligation of azide 7. HEK cells were transfected with constructs for expression of intracellular E2p-CFP or LAP-CFP together with one of the four LplA fusion constructs (FRB-eLplA, FRB-hLplA, mCherry-eLplA, or mCherry-hLplA). All cells were then grown in the presence of 500 μM azide **7** for ~ 24 hours before lysis. The azide was derivatized *in vitro* with biotin-alkyne in the presence of copper catalyst via the [3+2] cycloaddition and the ligated biotin was detected by streptavidin blotting. A negative control with E2p and LAP replaced by the alanine mutant of LAP (LAP(ala)) is also shown. eLplA stands for the enzyme expressed from the original *E. coli* lpla gene. hLplA stands for the enzyme expressed from a gene sequence optimized for expression in human cell lines.

We next performed the comparison of the specificity and yield of intracellular ligation of the unnatural substrate azide **7** by the different LplA constructs. Figure 8-10 shows that the difference in specificity observed between the eLplA and hLplA constructs for lipoic acid ligation disappeared in the case of azide **7** ligation. We observed that mCherry-eLplA, which was specific for lipoic acid ligation, was now ligating azide **7** to other mammalian proteins. FRB-eLplA was the only construct that did not display background labeling of cellular proteins, although its site-specific labeling signal of E2p was so low that it was hard to assess whether the signal/background was higher than for

the other LplA constructs. Both the low signal and low background were originated by the extremely low expression levels of intracellular FRB-eLplA. Later in this chapter, we will describe the imaging experiments that confirmed that FRB-eLplA was indeed specific for azide **7** ligation (see Figure 8-13).

Table 8-2 summarizes the results obtained from the systematic comparison of the different LplA constructs. The comparative study highlighted two origins for the observed background labeling: first, and most important, enzyme over expression, but also, de-coupling of the two steps involved in the mechanism of catalysis when going from the natural substrate lipoic acid to the unnatural azide **7** probe. First, the lipoic acid ligation results confirmed that background labeling was caused by LplA-catalyzed ligation of lipoic acid to many other cellular proteins, but not the endogenous lipoate-acceptor proteins. Second, we observed that the specificity of intracellular lipoic acid ligation is directly correlated to the enzyme expression levels. LplA looses its protein substrate specificity when highly over expressed. Third, transition from the ligation of lipoic acid to azide **7** also resulted in an additional loss of specificity. We hypothesized that this may be caused by a certain degree of de-coupling between the first step of azide **7** adenylation, and the second step of azide **7**-AMP transfer to the E2p or LAP substrates. If the enzyme displays higher affinity for the azide **7**-AMP than for azide **7,** the activated ester could then be released from the active site before transfer. Once in the cellular milieu, the azide **7** would react with any nucleophile, such as protein surface lysine residues, thus resulting in background labeling. Although plausible, this hypothesis does not explain the observation that, at very low expression levels, such as those obtained with the FRB-LplA construct or with tetracycline regulated mCherry-hLplA, site-specific ligation of azide **7** is possible. On the contrary, enzymatic de-coupling could also be caused by high enzyme affinity for the azide **7**-AMP ester, such that transfer to the protein substrate is not efficient and becomes the rate limiting step. Background labeling would thus be caused by azide **7**-AMP retention in the active site of LplA. In the lysate, this azide **7**-AMP would be released and could react with any surrounding protein. This theory would readily explain the fact that background staining using azide **7** is also higher at higher enzyme concentrations.

Finally, it should be highlighted that, when extremely low levels of enzyme expression were achieved by either endogenous protein regulation (FRB-eLplA) or by regulation with tetracyline (Tet-mCherry-hLplA), the enzyme recovered its ligation specificity for both lipoic acid and azide **7**. Although the results for the Tet-mCherry-hLplA are still preliminary, a comparison between the two constructs, suggested that, at similar expression levels, the improved folding and stability of hLplA over eLplA results in higher yields of specific labeling, thus generating higher signal/background ratios.

Table 8-2: Comparison of the specificity and extent of the lipoic acid and azide 7 by several LplA intracellular constructs.

LplA construct		Expression levels	Lipoic acid ligation		Azide 7 ligation	
Fusion	**Codon**		**Yield**	**Specific**	**Yield**	**Specific**
FLAG	*E. coli*	Low	ND	ND	Too low	ND
	human	Medium	High	No	Medium	No
mCherry	*E. coli*	Medium	Medium	Yes	Medium	No
	human	High	High	No	High	No
FRB	*E. coli*	Low	Medium	Yes	Low	Yes
	human	Medium	High	No	High	No
Tet System	*E. coli*	Regulated	Low	Yes	Too low	ND
	human	Regulated	Medium	Yes	Medium	Yes

The herein described western blot results were in all cases supported by the corresponding imaging data. The remaining sections in this chapter will show the imaging results confirming the site-specific ligation of the natural lipoic acid substrate, and of the unnatural substrates azide **7** and coumarins **1** and **2** .

Imaging the LplA-catalyzed site-specific incorporation of lipoic acid inside mammalian cells

We wished to confirm the specificity of mCherry-eLplA for the ligation of lipoic acid using the more demanding imaging readout. To more rigorously test the site-specificity of the ligation, we constructed a fusion of E2p-CFP to a C-terminal nuclear

localization signal (NLS), for nuclear expression of E2p. We transfected HEK cells with E2pCFP-NLS and the mCherry-eLplA plasmids, and, 24 hours after transfection, we added 500 μM lipoic to the cell growth media to initiate intracellular lipoic acid ligation. 24h later, cells were fixed and permeabilized, and ligated lipoic acid was detected by anti-lipoic acid immunostaining.

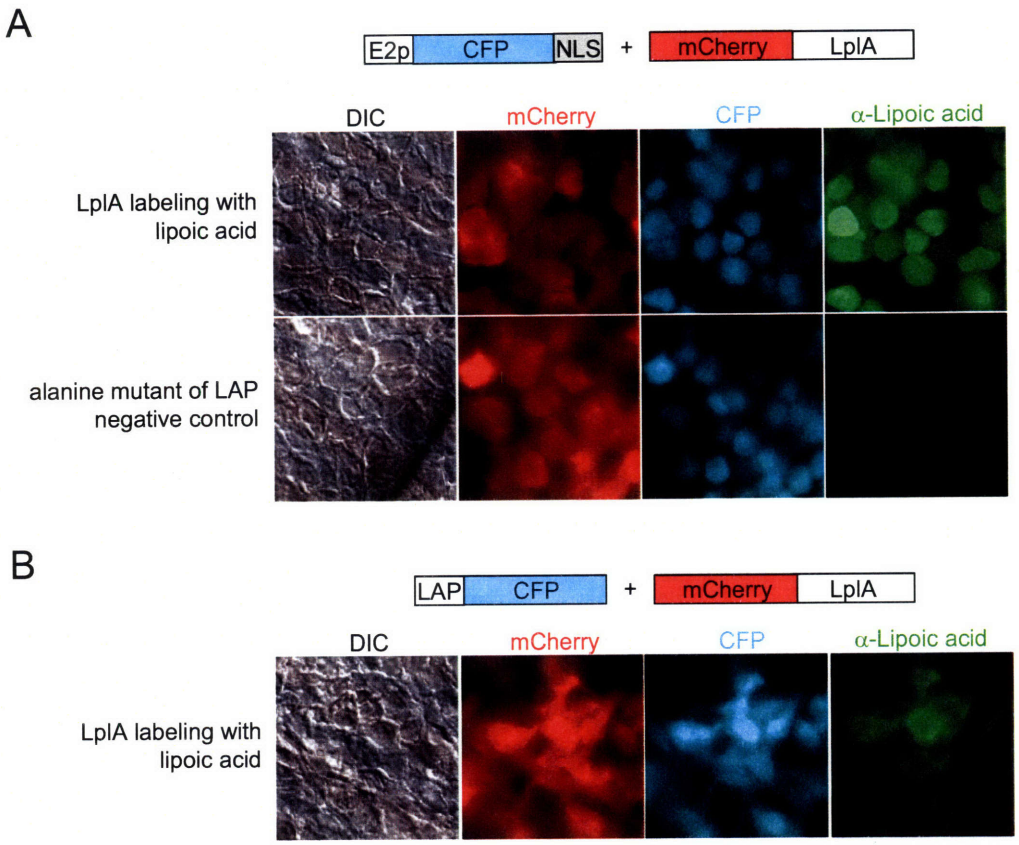


Figure 8-11: Imaging the site-specific intracellular ligation of lipoic acid to E2p and LAP. (**A**) HEK cells were co-transfected with nuclear targeted E2p-CFP and mCherry-eLplA. After 24 hours incubation with 500 μM lipoic acid, cells were fixed and permeabilized with formaldehyde/methanol, and stained with anti-lipoic acid antibody followed by anti-rabbit-Alexa Fluor 488. Anti-lipoic acid labeling is shown in green to the right of DIC (differential interference contrast), CFP (cyan), and mCherry (red) images. A negative control is shown where E2p-CFP was replaced by the alanine mutant of the LAP peptide. (**B**) Same as A, except that cells were co-transfected with mCherry-eLplA and LAP-CFP.

We were very happy to observe robust nuclear labeling on transfected cells, whereas untransfected cells or cells transfected with the LAP(ala) construct showed no labeling (Figure 8-11A). It must be noted that we used LAP(ala) as a negative control instead of E2p(ala) to avoid potential artifacts caused by enzyme inhibition. Furthermore,

the labeling was clearly nuclear, closely resembling the nuclear localization of E2p (CFP signal) and different from that of the predominantly cytoplasmic LplA (mCherry signal). The strong labeling signal observed for E2p encouraged us to test, for the first time, intracellular lipoic acid ligation onto the LAP peptide. We observed clear and reproducible labeling of cytosolic LAP-CFP (Figure 8-11B). As expected, the overall signal was reduced from that of lipoic acid, owing to the slower kinetics of LAP recognition by LplA.

Imaging the LplA-catalyzed site-specific incorporation of azide 7 inside mammalian cells

Next, we attempted to detect the ligation of azide **7** in live cells expressing FRB-eLplA. As previously described, western blot results suggested that expression of intracellular FRB-eLplA resulted in the site-specific ligation of azide **7** to E2p, albeit with a very low signal/background ratio. We transfected HEK cells with the FRB-eLplA construct together with E2p-CFP, and we performed the azide **7** labeling followed by detection with OCT-HDDA-CFDA as described above. Unfortunately, we could not reliably detect a significant site-specific fluorescein signal above background staining. Only under very forcing conditions, where azide **7** ligation was performed for 48 hours and the [3+2] cycloaddition reaction with OCT-HDDA-CFDA was allowed to proceed overnight, did we observe some specific labeling signal, albeit the signal/background ratio was < 1.5. We attempted to improve the signal/background ratio by using a faster cyclooctyne molecule with two electron withdrawing fluorines adjacent to the alkyne[18] (DIFO2-HDDA-CFDA), but we found that the increased reactivity was also accompanied by a marked increase in its non-specific reactivity toward other cellular nucleophiles.

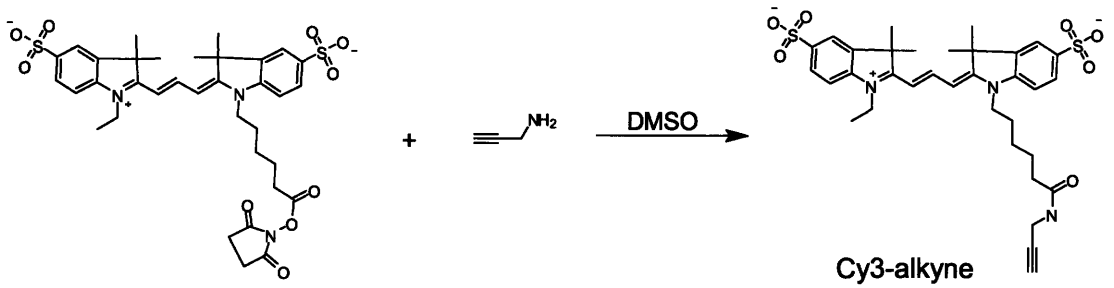


Figure 8-12: Synthesis of Cy3-alkyne for click chemistry detection of azide 7 ligation.

To undoubtedly demonstrate the site-specificity of azide **7** ligation by FRB-eLplA, we decided to increase the labeling signal by employing the faster Cu(I)-mediated [3+2] cycloaddition. We synthesized a terminal alkyne conjugated to the Cy3 fluorophore by direct coupling of the *N*-hydroxysuccinimidyl ester of Cy3 to propargylamine (Figure 8-12). Because Cu(I) is toxic to mammalian cells, we performed azide **7** ligation in live cells as above, but we then fixed and permeabilized the cells before detecting the azide by click chemistry with Cy3-alkyne. We were very happy to see clear and specific detection of azide **7** ligation to E2p (Figure 8-13A). As predicted by the Western blot results, the signal obtained for ligation to the LAP peptide was barely detectable (Figure 8-13B).

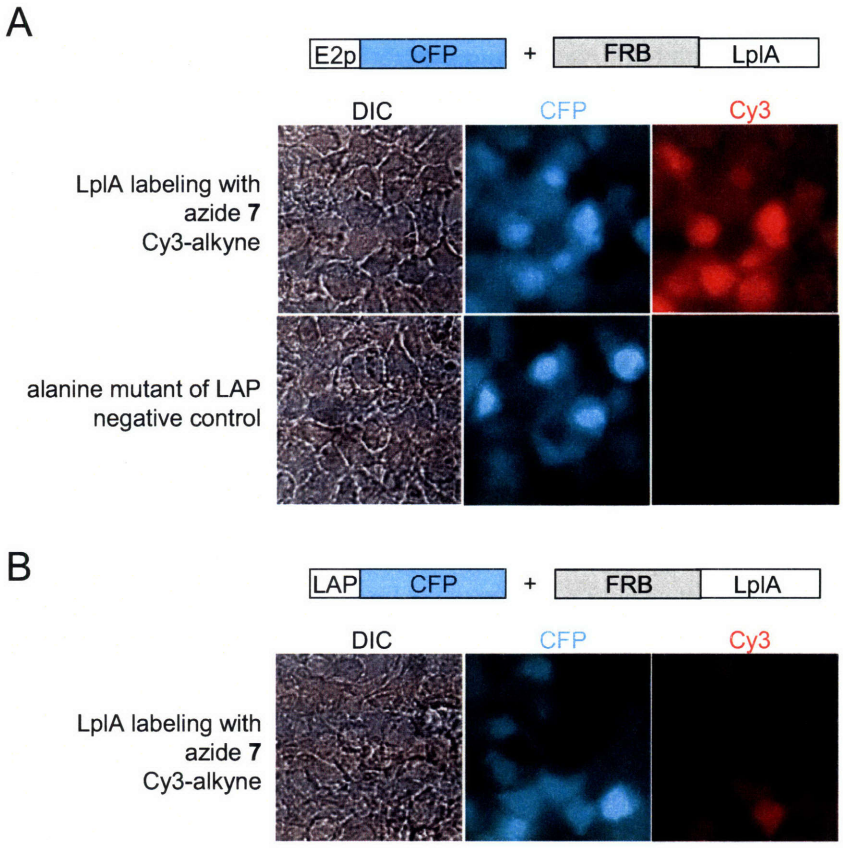


Figure 8-13: Fixed-cell detection of intracellular ligation of azide 7 via click chemistry. (**A**) E2p-CFP and FRB-eLplA fusions were expressed in the cytosol of HEK cells. Intracellular E2p was first labeled with azide **7** by the expressed eLplA before cells were fixed and permeabilized with formaldehyde/methanol, and the introduced azide was detected by click chemistry reaction with Cy3-alkyne. Fixed-cell images of the introduced Cy3 fluorophore are shown to the right of the CFP and DIC images, which highlight the transfected cells. A negative control is shown where E2p-CFP was replaced by the alanine mutant of the LAP peptide. (**B**) Same as A, except that cells were co-transfected with mCherry-eLplA and LAP-CFP.

These results were extremely exciting because they showed, for the first time, that specific ligation of azide **7** by intracellular LplA is possible. Since these results were obtained, Dr. Shah in our lab has been working on using the tetracycline-regulated system to increase the signal/background ratio such that live-cell detection using OCT-HDDA-CFDA is possible. He recently reported that, under forcing conditions like the ones previously described for FRB-eLplA (i.e., 48 hours treatment with azide **7** followed by overnight [3+2] cycloaddition with OCT-HDDA-CFDA in the presence of probenecid), specific ligation of azide **7** can be detected in live cells with a decent signal/background ratio. He has also shown that decreasing the concentration of azide **7** from 500 μM to ~ 100–200 μM further improves the signal, probably owing to a more efficient wash out of excess probe from cells. Importantly, targeting of E2p to the nucleus resulted in a clearly nuclear-localized fluorescein signal, thus reinforcing the fact that the labeling is site-specific.

Direct fluorophore ligation to intracellular proteins in live mammalian cells

Finally, we reasoned that, if the rate limiting step that was preventing the reliable detection of azide **7** ligation in live cells was the poor kinetics of the strained-promoted cycloaddition reaction, perhaps the problem could be circumvented by employing a one-step labeling protocol, using the direct ligation of a coumarin fluorophore. The labeling procedure for direct fluorophore ligation is much simpler, as it only requires co-expression of the LplA W37I mutant together with E2p, followed by incubation with a membrane-permeant coumarin derivative (Figure 8-14A). Dr. Hemanta Baruah synthesized several protected versions of coumarins **1** and **2** (Figure 7-13). Although the use of coumarin **1** was preferred for intracellular applications because of its lower pKa (i.e., higher percentage of fluorescent molecules at physiological pH), protection of the hydroxyl group proved intractable, owing as well to the low pKa. Protection with a bulky tertbutyl group rendered the molecule stable in aqueous solutions, but we observed extremely slow deprotection kinetics inside cells, which lengthened the required labeling time. Thus, we opted to use an acetate-protected version of coumarin **2**, which was relatively stable in the cell loading media (coumarin **2**-acetate, Figure 8-14B).

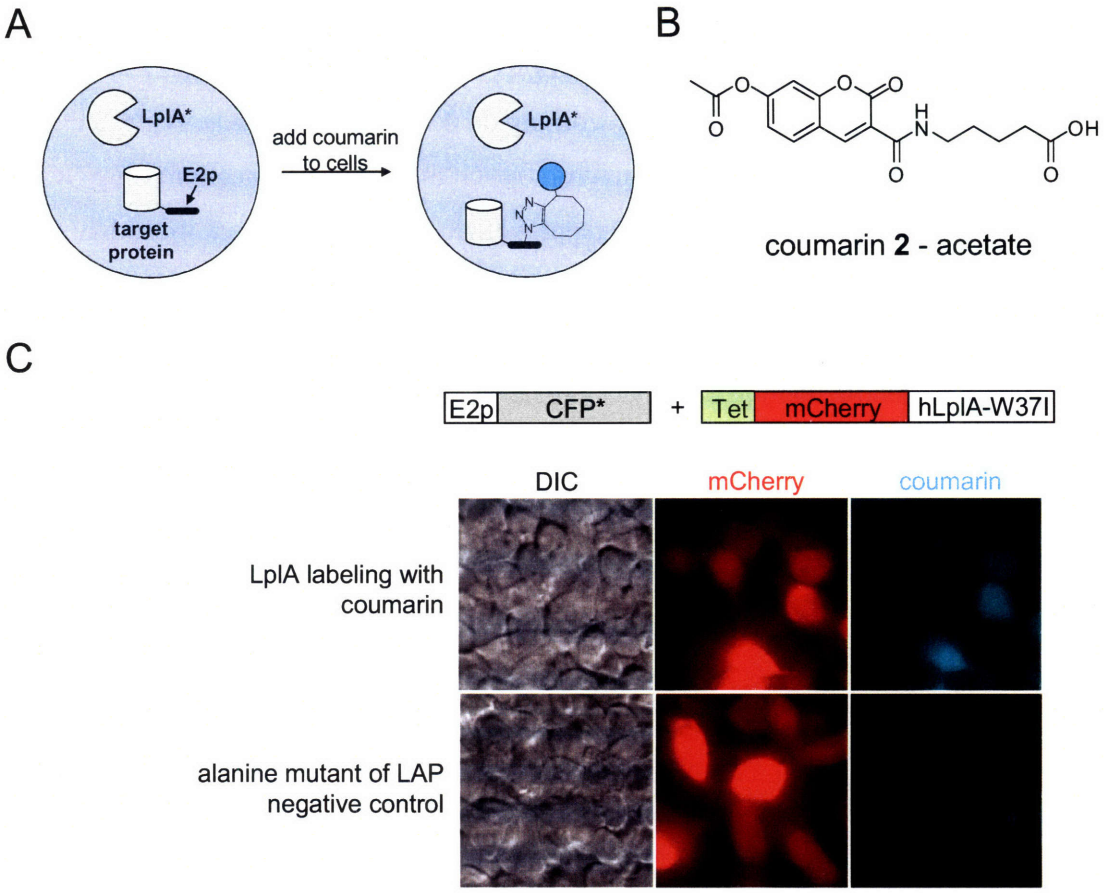


Figure 8-14: One-step fluorophore tagging of intracellular proteins in live mammalian cells. (**A**) Labeling procedure. Cells are first co-transfected with plasmids encoding a LAP- or E2p-fused protein and a mutant of the lipoic acid ligase enzyme (LplA-W37I). Labeling is performed in single step by addition of a membrane-permeant coumarin probe. (**B**) Structure of coumarin **2**-acetate (6,8-fluoro-7-hydroxycoumarin acetate). (**C**) Live cell detection of intracellular coumarin ligation. HEK cells were co-transfected with a fusion of E2p to a non-fluorescence mutant of CFP (W66L) (E2p-CFP*) and a tetracycline-regulated fusion of the W37I mutant of LplA to mCherry (Tet-mCherry-hLplA-W37I). Protein expression was induced 24 hours after transfection by addition of 0.5 μg/mL tetracycline. Intracellular ligation was performed 24 hours later, by incubation with 500 μM coumarin **2**-acetate for 2 hours at 32 °C. Live cell images of the introduced coumarin fluorophore are shown to the right of the mCherry and DIC images, which highlight the transfected cells. A negative control is shown where E2p-CFP was replaced by the alanine mutant of the LAP peptide.

To perform live cell labeling, we transfected HEK cells with the non-fluorescent E2p-CFP* and the tetracycline-regulated mCherry-hLplA(W37I) constructs. Enzyme expression was induced by addition of tetracycline, and labeling was performed 16–24 hours later by incubation with 500 μM coumarin **2**-acetate for 2 hours at 32 °C. The results are shown in Figure 8-14C. We were delighted to observe specific ligation of coumarin to cytosolic E2p, as these results represented the first time that LplA-catalyzed

probe ligation to intracellular proteins was detected in live cells. Unfortunately, as previously described, the W37I mutant of LplA does not recognize the LAP peptide as efficiently as the wild-type enzyme, and thus we have not been able to detect coumarin ligation to the peptide substrate yet.

Dr. Hemanta Baruah, Tao Uttamapinant, Katie White, and Dr. Sujiet Puthenveetil, in our lab, are actively working on synthesizing better coumarin probes, finding faster mutants of LplA, and evolving and testing improved peptide substrates. For example, Sujiet Puthenveetil has found, using yeast display evolution, a peptide that is recognized by LplA with similar kinetics to E2p *in vitro*. Furthermore, Katie White has just recently shown that intracellular lipoylation of the improved peptide yields a signal that is comparable to that obtained for the E2p protein substrate. We expect that a combination of all the above efforts will ultimately allow the direct visualization of intracellular tagging of LAP-fused proteins in living cells.

Conclusions

We have shown that the site-specific intracellular ligation of unnatural probes by lipoic acid ligase is possible and that this ligation can be used to fluorescently label E2p-tagged intracellular proteins. As expected, adaptation of lipoic acid ligase for intracellular protein labeling proved a lot more challenging than its application to label cell-surface receptors. We found that LplA misfolded and aggregated when expressed inside mammalian cells, and that this aggregation, together with enzyme overexpression, resulted in the loss of specificity for its protein substrate. Strikingly, high levels of enzyme overexpression inside cells resulted in LplA-catalyzed ligation of lipoic acid and azide **7** to myriad cellular proteins. Through an iterative process, we were able to recover the lost enzyme specificity. We decreased enzyme aggregation by fusion to the mCherry fluorescent protein, we improved enzyme folding by using a gene sequence optimized for mammalian expression, and we introduced protein regulation (i.e., turnover) by fusion to the human FRB protein. Expression of LplA under the regulation of a tetracyline operator further enhanced our capability to tightly control enzyme concentration inside cells. After

implementation of all these changes, we were able to show specific ligation of lipoic acid both by western blot and by immunostaining. We also demonstrated the site-specificity of the azide **7** ligation by western blot and by performing click chemistry on fixed cells. Unfortunately, the toxicity of the Cu(I)-mediated [3+2] cycloaddition and the poor kinetics of the strained-promoted counterpart, prevented us from visualizing azide **7** ligation in live cells. We nonetheless circumvented this problem by eliminating the need for bioorthogonal ligation. The discovery that a W37I mutant of LplA could efficiently ligate coumarin derivatives to E2p-tagged proteins enabled, for the first time, visualization of LplA-catalyzed probe ligation in living cells.

Finally, it must be noted that fixed-cell detection of azide **7** ligation using click chemistry can already be applied for super-resolution imaging of intracellular proteins using a recently developed imaging technique called STORM (stochastic optical reconstruction microscopy)[19]. In STORM imaging, proteins labeled with cyanine fluorophores can be localized inside cells with the astounding precision of ~ 20 nm, a ~ 12-fold improvement over conventional wide-field or confocal microscopy[20,21]. Cyanine dyes are however targeted for STORM imaging using antibodies, which add ~ 50 nm uncertainty in the localization precision, thus partially defeating the purpose of super-resolution imaging. Therefore, in order to optimally utilize the intrinsic imaging resolution of the super-resolution techniques, new strategies to target the cyanine dyes without significantly increasing the size of the tag and with high labeling efficiency are needed. We have now demonstrated that LplA labeling can do so. Furthermore, Tao Uttamapinant has recently demonstrated that performing both labeling steps (enzymatic ligation of azide **7** and click chemistry with Cy5-alkyne) after cell fixation, allows the fluorescent labeling of LAP-tagged intracellular proteins with a signal/background ratio of ~ 4.

Experimental

General synthetic methods

Reagents were purchased from Sigma-Aldrich, Alfa Aesar, TCI America, Invitrogen, or GE Healthcare and used without further purification. Analytical thin layer chromatography (TLC) was performed using 0.25 mm silica gel 60 F_{254} plates and visualized with ninhydrin or bromocresol. Flash column chromatography was carried out using silica gel (ICN SiliTech 32-63D). Mass spectra were recorded on an Applied Biosystems 200 QTRAP Mass Spectrometer using electrospray ionization. HPLC was performed on a Varian Prostar Instrument equipped with an autosampler and photo-diode-array detector. For analytical HPLC, a reverse-phase 250 × 4.6 mm Microsorb-MV 300 C18 column was used. For preparative HPLC, a reverse-phase 250 × 10 mm Microsorb-MV 100 C18 column was used. Chromatograms were recorded at 210 nm unless otherwise noted.

Synthesis of membrane-permeant OCT-fluorophore conjugates

Synthesis of OCT-PFP (2)

To a solution of OCT acid **1**[22] (17 mg, 65 μmol) in dry dichloromethane (DCM, 750 μL) was added triethylamine (TEA, 18 μL, 130 μmol). The mixture was stirred for 5 minutes at ambient temperature. Pentafluorophenyl trifluoroacetate (PFP-TFA, 22 μL, 130 μmol) was added slowly to the reaction mixture over 3 minutes and the reaction was allowed to proceed for 3 hours. The reaction mixture was concentrated in *vacuo*, then purified by silica chromatography (10% ethyl acetate in hexane) to afford OCT-PFP **2** as a colorless solid (25 mg, 56 μmol, 90%).

Synthesis of OCT-HDDA (3)

Freshly prepared OCT-PFP **2** solid (100mg, 234 μmol) was immediately dissolved in dry DCM (4 mL), and then combined with TEA (65 μL, 468 μmol) and

N,N'-dimethyl-1,6-diaminohexane (HDDA, 169 mg, 1170 μmol). The reaction mixture was allowed to stir overnight at ambient temperature. The crude mixture was purified on a silica flash column (10% methanol in DCM). The purified product was concentrated under reduced pressure to yield OCT-HDDA **3** (50 mg, 129 μmol, 55%). ESI-MS calculated for $[M + H]^+$: 387.27; observed 387.36.

Synthesis of OCT-EDA (4)

Freshly prepared OCT-PFP **2** solid (30 mg, 70 μmol) was immediately dissolved in dry DCM (2 mL), and then combined with TEA (29 μL, 210 μmol) and ethylene diamine (EDA, 38 μL, 560 μmol). The reaction mixture was allowed to stir overnight at ambient temperature. The crude product was purified by HPLC on a reverse phase column. The following conditions were used for elution: 30–100% acetonitrile in water over 30 minutes; flow rate 10.0 mL/minute. The purified product was concentrated under reduced pressure to yield OCT-EDA **4** (50 mg, 129 μmol, 55%). ESI-MS calculated for $[M + H]^+$: 303.18; observed 303.30.

Synthesis of HDDA-CF (5)

To a solution of the *N*-hydroxysuccinimidyl ester of 5/6-carboxyfluorescein (50 mg, 105 μmoles) in anhydrous DMSO (300 μL) was added TEA (41 μL, 300 μmoles). The mixture was stirred for 1 minute before *N,N'*-dimethyl-1,6-diaminohexane (72 mg, 500 μmol) was added to the mixture. The reaction was allowed to proceed overnight at 30 °C. The crude product was purified by HPLC on a reverse phase column. The following conditions were used for elution: 30–100% acetonitrile in water over 30 minutes; flow rate 10.0 mL/minute. Solvent was removed in *vacuo* to afford HDDA-CF **5**. ESI-MS calculated for $[M + H]^+$: 503.21; observed 503.16.

Synthesis of OCT-HDDA-CFDA (6)

To a solution of HDDA-CF **5** (15 mg, 29 μmol) in anhydrous DMF (100 μL) was added TEA (8 μL, 58 μmol). OCT-PFP **2** (10 mg, 23 μmol) was then added to the

mixture and the reaction was allowed to stir overnight at ambient temperature. Thereafter, acetic anhydride (17 μL, 174 μmol) was added and the reaction was allowed to stir for 5 minutes at ambient temperature. The solution turned from yellow to colorless indicating formation of the desired product. The crude product was purified by HPLC on a reverse phase column. The following conditions were used for elution: 30–100% acetonitrile in water over 30 minutes; flow rate 10.0 mL/minute. Solvent was removed in *vacuo* to afford OCT-HDDA-CFDA **6**. ESI-MS calculated for $[M - H]^-$: 829.34; observed 829.34.

Synthesis of OCT-HDDA-BODIPY TR (7)

To a solution of OCT-HDDA **3** (1 mg, 2.5 μmol) in anhydrous DMSO (100 μL) was added TEA (0.7 μL, 5 μmol) and the solution was stirred for 2 minutes. The 4-sulfotetrafluorophenyl (STP) ester of BODIPY TR (1 mg, 1.25 μmol) was then added and the reaction was allowed to stir for 4 hours at ambient temperature. The crude product was purified by HPLC on a reverse phase column. Chromatograms were recorded both at 210 nm and 550 nm. The following conditions were used for elution: 30–80% acetonitrile in water over 30 minutes; flow rate 5.0 mL/minute. Solvent was removed in *vacuo* to afford OCT-HDDA-BODIPY TR (**7**). Estimated yield 35%. ESI-MS calculated for $[M + H]^+$: 906.43; observed 906.36.

*Synthesis of OCT-HDDA-BODIPY FL (**8**)*

To a solution of OCT-HDDA **3** (1 mg, 2.5 μmol) in anhydrous DMSO (100 μL) was added TEA (0.7 μL, 5 μmol) and the solution was stirred for 2 minutes. The 4-sulfotetrafluorophenyl (STP) ester of BODIPY FL (1 mg, 1.8 μmol) was then added and the reaction was allowed to stir for 4 hours at ambient temperature. The crude product was purified by HPLC on a reverse phase column. The following conditions were used for elution: 30–80% acetonitrile in water over 30 minutes; flow rate 5.0 mL/minute. Solvent was removed in *vacuo* to afford OCT-HDDA-BODIPY FL **8**. Estimated yield 50%. ESI-MS calculated for $[M - H]^-$: 659.38; observed 659.34.

Synthesis of OCT-HDDA-SNARF®-1 (9)

To a solution of OCT-HDDA **3** (0.2 mg, 0.5 μmol) in anhydrous DMSO (100 μL) was added TEA (0.2 μL, 1.5 μmol) and the solution was stirred for 2 minutes. The *N*-hydroxysuccinimidyl ester of SNARF®-1 acetate (0.3 mg, 0.5 μmol) was then added and the reaction was allowed to stir for 4 hours at ambient temperature. The crude product was used without further purification. ESI-MS calculated for $[M + H]^+$: 864.39; observed 864.42.

Synthesis of OCT-EDA-CFDA (10)

To a solution of OCT-EDA **4** (1.1 mg, 3.6 μmol) in anhydrous DMSO (100 μL) was added diisopropylethylamine (DIEA, 2 μL, 10.8 μmol) and the solution was stirred for 2 minutes. The *N*-hydroxysuccinimidyl ester of 5/6-carboxyfluorescein diacetate (2 mg, 3.6 μmol) was then added and the reaction was allowed to stir for 4 hours at ambient temperature. During the course of the reaction, the acetate groups hydrolyzed. Re-protection was afforded by addition of excess acetic anhydride. The crude product was purified by HPLC on a reverse phase column. The following conditions were used for elution: 30–80% acetonitrile in water over 30 minutes; flow rate 5.0 mL/minute. Solvent was removed in *vacuo* to afford OCT-EDA-CFDA **10**. Estimated yield 58%. ESI-MS calculated for $[M - H]^-$: 701.24; observed 701.10.

Synthesis of OCT-EDA-BODIPY FL (11)

To a solution of OCT-EDA **4** (1.1 mg, 3.6 μmol) in anhydrous DMSO (100 μL) was added TEA (2 μL, 10.8 μmol) and the solution was stirred for 2 minutes. The 4-sulfotetrafluorophenyl (STP) ester of BODIPY FL (1 mg, 1.8 μmol) was then added and the reaction was allowed to stir for 4 hours at ambient temperature. The crude product was purified by HPLC on a reverse phase column. The following conditions were used for elution: 30–80% acetonitrile in water over 30 minutes; flow rate 5.0 mL/minute. Solvent was removed in *vacuo* to afford OCT-EDA-BODIPY FL **11**. Estimated yield 44%. ESI-MS calculated for $[M + Na]^+$: 599.19; observed 599.58.

Synthesis of membrane-permeant DIFO2-CFDA

Synthesis of DIFO2-HDDA-CFDA (14)

To a solution of DIFO2 acid **12**[22] (10 mg, 36 μmol) in dry DCM (750 μL) was added TEA (10 μL, 70 μmol). The mixture was stirred for 5 minutes at ambient temperature. PFP-TFA (19 μL, 108 μmol) was added slowly to the reaction mixture over 3 minutes and the reaction was allowed to proceed for 3 hours. The reaction mixture was concentrated in *vacuo*, then purified by preparatory TLC (10% ethyl acetate in hexane) to afford DIFO2-PFP **13** as a colorless solid (2 mg, 4.5 μmol, 12%).

To a solution of HDDA-CF **5** (5 mg, 9.6 μmol) in anhydrous DMF (100 μL) was added TEA (2.7 μL, 29 μmol). DIFO2-PFP **13** (2 mg, 4.5 μmol) was then added to the mixture and the reaction was allowed to stir overnight at ambient temperature. Thereafter, acetic anhydride (2.6 μL, 27 μmol) was added and the reaction was allowed to stir for 5 minutes at ambient temperature. The solution turned from yellow to colorless indicating formation of the desired product. The crude product was purified by HPLC on a reverse phase column. The following conditions were used for elution: 30–100% acetonitrile in water over 30 minutes; flow rate 10.0 mL/minute. Solvent was removed in *vacuo* to afford DIFO2-HDDA-CFDA **14**. ESI-MS calculated for $[M - H]^-$: 847.33; observed 847.68.

Synthesis of Cy3-alkyne

To 18 μL of a solution of the *N*-hydroxysuccinimidyl ester of Cy3 (1 mg, 1.3 μmol) in anhydrous DMSO (80 μL) was added propargylamine (214 μg, 0.3 μL, 3.9 μmol) and the reaction mixture was stirred at ambient temperature for 3 hours. The crude product was purified by HPLC on a reverse phase column. Chromatograms were recorded both at 210 nm and 550 nm. The following conditions were used for elution: 10–50% acetonitrile in water over 30 minutes; flow rate 5.0 mL/minute. Solvent was removed in *vacuo* to afford Cy3-alkyne. Estimated yield 70–80%. ESI-MS calculated for $[M - H]^-$: 666.23; observed 666.36.

Synthesis of a membrane-permeant coumarin probe (coumarin 2-acetate)

To a solution of coumarin **2** (5 mg, 16 μmol) in anhydrous DMSO (100 μL) was added TEA (4.4 μL, 32 μmol) and excess acetic anhydride and the reaction mixture was stirred for 1 hour at ambient temperature in the dark. The crude product was purified by HPLC on a reverse phase column. The following conditions were used for elution: 30–80% acetonitrile in water over 30 minutes; flow rate 10.0 mL/minute. Solvent was removed in *vacuo* to afford coumarin **2**-acetate. Estimated yield 37%. ESI-MS calculated for $[M - H]^-$: 346.10; observed 346.08.

Cloning of HA-E2p for cytoplasmic mammalian expression

The gene encoding for a 9kD hybrid lipoyl domain derived from the second subunit of the *E. coli* pyruvate dehydrogenase[7] was inserted between the *BamHI* and *EcoRI* sites of a modified form of the pcDNA3 vector (Invitrogen). The gene was PCR-amplified from the original pGS331 plasmid[7], a gift from John Cronan, with primers 5′ AAAAG GAT CCA TAT CCG TAC GAC GTA CCA GAC TAC GCA ATG GCT ATC GAA ATC AAA GTA CCG G and 5′ TTTT GAA TTC TTA CGC AGG AGC TGC CGC AGG CG to introduce an HA tag (YPYDVPDYA) and a *BamHI* site at the N-terminus, and a stop codon and *EcoRI* site at the C-terminus. The resulting PCR product was digested with *BamHI* and *EcoRI* and ligated in-frame to *BamHI/EcoRI* digested pcDNA3, to give pcDNA3-HA-E2p.

Cloning of E2p-CFP for cytoplasmic mammalian expression

The gene encoding for a 9kD hybrid lipoyl domain derived from the second subunit of the *E. coli* pyruvate dehydrogenase[7] was inserted between the *NheI* and *BamHI* and sites of a modified form of the pcDNA3 vector (Invitrogen). The vector contained the CFP gene between the *BamHI* and *EcoRI* sites[3].The E2p gene was PCR-amplified from the original pGS331 plasmid[7], a gift from John Cronan, with primers 5′ AAAA GCT AGC ATG GCT ATC GAA ATC AAA GTA CCG G and 5′ TTTT GG ATC CTT CGC AGG AGC TGC CGC AGG CG. The resulting PCR product was digested with *NheI* and *BamHI* and ligated in-frame to *NheI/BamHI* digested pcDNA3, to give pcDNA3-E2p-CFP. To create the

E2p(Ala)-CFP mutant, we performed QuikChange with the primer 5′ C ACC GTA GAA GGC GAC GCA GCT TCT ATG GAA GTT CCG and its reverse complement. To create a non-fluorescent E2p-CFP (pcDNA3-E2p-CFP*), Trp66 was mutated to leucine by QuickChange with primer 5′ G ACC ACC CTG ACC CTG GGC GTG CAG TGC TTC and its reverse complement.

Cloning of E2p-CFP-NLS for nuclear mammalian expression

The E2p-CFP gene was PCR-amplified from the corresponding pCDNA3 plasmid with the primers: 5′ AAAA GCT AGC ATG GCT ATC GAA ATC AAA GTA CCG G and 5′ TTTT GAA TTC CTT TAC ACC TTG CGC TTC TTC TTG GGC ACC TTG CGC TTC TTC TTG GGC ACC TTG CGC TTC TTC TTG GGG CCG CCG GAG GAC TCC TTG TAC to introduce an *NheI* site at the N-terminus, and three tandemly repeated nuclear localization signals (NLS, protein sequence PKKKRKVPKKKRKVPKKKRKV) and an *EcoRI* site at the C-terminus. The PCR product was ligated into the *NheI* and *EcoRI* sites of pcDNA3 (Invitrogen). To create the E2p(Ala)-CFP-NLS mutant, we performed QuikChange with the primer 5′ C ACC GTA GAA GGC GAC GCA GCT TCT ATG GAA GTT CCG and its reverse complement.

Cloning of LAP-CFP for cytoplasmic mammalian expression

Cloning of this construct was described in Chapter 6.

Cloning of FLAG-LplA for cytoplasmic mammalian expression

The original *E. coli* gene encoding for *E. coli* lipoic acid ligase (LplA) was inserted between the *BamHI* and *EcoRI* sites of a modified form of the pcDNA3 vector (Invitrogen). The gene was PCR-amplified from the original pYFJ16 plasmid, a gift from John Cronan, with primers 5′ AAAA G GAT CCA GAC TAC AAG GAT GAC GAC GAT AAG TCC ACA TTA CGC CTG CTC ATC and 5′ TTTT GAA TTC TTA CAG CCC CCG CCA TCC ATG to introduce a FLAG tag (DYKDDDDK) and a *BamHI* site at the N-terminus, and a stop codon and *EcoRI* site at the C-terminus. The resulting PCR product was digested with *BamHI* and *EcoRI* and ligated in-frame to *BamHI/EcoRI* digested pcDNA3, to give

pcDNA3-FLAG-LplA. To create the catalytically inactive mutant, Lys133 was mutated to arginine by QuikChange using 5′ CGT CGA AGG CGA CCG CAG AGT CTC AGG CTC GGC CTA T and its reverse complement.

Cloning of mCherry-LplA for cytoplasmic mammalian expression

A plasmid containing the gene encoding for the fluorescence protein mCherry was a gift from Roger Tsien. The plasmid contained the mCherry gene inserted between the *NheI* and *BamHI* of the pcDNA3 vector (Invitrogen). The LplA gene was digested out of pcDNA3-FLAG-LplA with *BamHI* and *EcoRI* and ligated in-frame to *BamHI/EcoRI* digested pcDNA3-mCherry, to give pcDNA3-mCherry-LplA (with the FLAG epitope in between both proteins). To create the catalytically inactive mutant, Lys133 was mutated to arginine by QuikChange using 5′ CGT CGA AGG CGA CCG CAG AGT CTC AGG CTC GGC CTA T and its reverse complement.

Cloning of FRB-LplA for cytoplasmic mammalian expression

The human FRB gene (a gift from Paul Clemons) was PCR-amplified using the primers 5′ AAAA GCT AGC TAT CCG TAC GAC GTA CCA GAC TAC GCA ATG TGG CAT GAA GGC CTG and 5′ TTTT GGA TCC TTG CCC GAG CCC GAG GTC GAG CCC GAG CCC TTT GAG ATT CGT CGG AA to introduce an *NheI* site and an HA tag (YPYDVPDYA) at the N-terminus, and a 9-amino acid linker (GSGSTSGSG) and a *BamHI* site at the C-terminus. The PCR product was digested with *NheI* and *BamHI* and ligated in-frame to *NheI/BamHI* digested pcDNA3-mCherry-LplA, to give pcDNA3-FRB-LplA (with the FLAG epitope in between both proteins).

Cloning of FLAG-hLplA (codon optimized) for cytoplasmic mammalian expression

The codon-optimized gene for mammalian expression of *E. coli* lipoic acid ligase (LplA) was inserted between the *BamHI* and *EcoRI* sites of a modified form of the pcDNA3 vector (Invitrogen). The synthetic gene corresponding to a DNA sequence optimized for expression of the LplA enzyme in human cell lines was purchased from the

GeneScript Corporation. The gene was PCR-amplified with primers 5′ AAAA G GAT CCA GAC TAC AAG GAT GAC GAC GAT AAG ATG AGC ACC CTG AGA CTG CTG ATC and 5′ TTTT GAA TTC TTA TCT CAC GGC GCC GGC CAT CCA G to introduce a FLAG tag (DYKDDDDK) and a *BamHI* site at the N-terminus, and a stop codon and *EcoRI* site at the C-terminus. The resulting PCR product was digested with *BamHI* and *EcoRI* and ligated in-frame to *BamHI/EcoRI* digested pcDNA3, to give pcDNA3-FLAG-hLplA (where hLplA stands for 'humanized' LplA).

Cloning of mCherry-hLplA (codon optimized) for cytoplasmic mammalian expression

To construct the pcDNA3-mCherry-hLplA plasmid, the PCR product obtained above was ligated in-frame to *BamHI/EcoRI* digested pcDNA3-mCherry. To create the catalytically inactive mutant, Lys133 was mutated to arginine by QuikChange using 5′ GAG GGC GAC AGA AGG GTG AGC GGC AGC and its reverse complement.

Cloning of FRB-hLplA (codon optimized) for cytoplasmic mammalian expression

The human FRB gene was PCR-amplified as above and ligated in-frame to *NheI/BamHI* digested pcDNA3-mCherry-hLplA, to give pcDNA3-FRB-hLplA (with the FLAG epitope in between both proteins).

Cloning of mCherry-hLplA (codon optimized) for cytoplasmic mammalian expression under the regulation of a tetracycline operator

The pcDNA4-TO-mCherry-hLplA construct was created by direct digestion from the corresponding pcDNA3 plasmid followed by ligation into the *HindIII* and *EcoRI* sites of pcDNA4-Tet. To create the LplA mutant that ligates the coumarin probes, Trp37 was mutated to isoleucine by QuikChange using 5′ CCA GAG AGT GCT GTT CCT GAT CAG AAA CGC CGA CAC CGT G and its reverse complement.

Detection of intracellular azide 7 ligation to E2p by native-gel mobility shift assay

Human cervical carcinoma (HeLa) cells were transfected with the pcDNA3- HA-E2p and pcDNA3-FLAG-LplA plasmids using Lipofectamine 2000 (0.8 μg of each DNA per well of a 24-well plate). From the moment of transfection, cells were grown in the presence of 750 μM azide **7**. Lysates were generated 48 hours later by hypotonic lysis to minimize protease release, as follows. Cells were lifted from the plates, concentrated by centrifugation, and resuspended in 1 mM HEPES pH 7.5, 5 mM magnesium chloride, 1 mM phenylmethylsulphonyl fluoride, and protease inhibitor cocktail (Calbiochem). After incubation at 4 °C for 10 minutes, the cells were lysed by vigorous vortexing for 2 minutes at 21 °C. Crude lysate was clarified by centrifugation, and either analyzed immediately or stored at –80 °C. When indicated, lysate was further labeled *in vitro* by incubating at 30 °C for 10 hours with 25 mM sodium phosphate pH 7.0, 1 μM LplA, 250 μM azide **7**, 1 mM ATP, and 4 mM magnesium acetate. Thereafter, cell extracts were analyzed by 20% native PAGE followed by Western blotting with rabbit anti-HA antibody (Rockland Immunochemicals, 1:1000 dilution) followed by anti-rabbit-horseradish peroxidase conjugate (Pierce, 1:2000 dilution).

Detection of intracellular azide 7 ligation to E2p via [3+2] cycloaddition with biotin-alkyne

Human embryonic kidney (HEK) 293T cells were transfected with the pcDNA3-E2p-CFP and pcDNA3-FLAG-LplA (Figure 8-3) or pcDNA3-mCherry/FRB-e/hLplA (Figure 8-10) plasmids using Lipofectamine 2000 (0.8 μg of each DNA per well of a 24-well plate). 0–24 hours after transfection, 500–750 μM azide **7** was added to the cell growth media (DMEM with 10% FBS, 50 units/mL penicillin, and 50 μg/mL streptomycin). Lysates were generated 24–48 hours later by hypotonic lysis to minimize protease release, as follows. Cells were lifted from the plates, concentrated by centrifugation, and resuspended in 1 mM HEPES pH 7.5, 5 mM magnesium chloride, 1 mM phenylmethylsulphonyl fluoride, and protease inhibitor cocktail (Calbiochem). After incubation at 4 °C for 10 minutes, the cells were lysed by vigorous vortexing for 2 minutes at 21 °C. Crude lysate was clarified by centrifugation, and stored at –80 °C.

Thereafter, azide derivatization via the [3+2] cycloaddition was performed by adding biotin-alkyne (BALK, prepared by Irwin Chen as described in his thesis[8]) to a final concentration of 500 μM in the presence of 1 mM $CuSO_4$, 2 mM tris(triazolyl)amine ligand (Sharpless ligand), and 4 mM tris(2-carboxyethyl)phosphine (TCEP) and incubating at 30 °C for 16 hours. Each reaction sample was then divided into thirds. The first third was analyzed by 12% SDS-PAGE followed by Western blotting with streptavidin-peroxidase conjugate (Pierce, 1:3000 dilution). The second sample was analyzed by 12% SDS-PAGE followed by Coomassie staining. The last third was analyzed by 12% SDS-PAGE without boiling the samples, in order to prevent unfolding of CFP, and in-gel fluorescence was visualized on a Storm 860 instrument (Amersham).

Assay of membrane permeability of the OCT-fluorophore conjugates

Stocks of all OCT-fluorophores conjugates were prepared in dry DMSO at 200x the final concentration, to ensure that the concentration of DMSO in the final cell-loading solution was kept constant for all probes at 1%. The cell-loading solution was prepared by dilution of 100x OCT-fluorophore stock to 1x in regular cell growth media (DMEM with 10% FBS, 50 units/mL penicillin, and 50 μg/mL streptomycin) containing 0.1% Pluronic F-127 (Invitrogen) and 1% DMSO. HEK 293T or HeLa cells were incubated with cell loading for a given period of time ranging from 5–60 minutes at a range of temperatures 16–37 °C. Thereafter, cells were rinsed three with cell growth media and imaged immediately. After imaging, cells were incubated further at 21–37 °C to allow the corresponding OCT-fluorophore to diffuse or be pumped-out of the cells. After 15–60 minutes, cells were imaged again. In all cases, cells were imaged in DPBS on a Zeiss Axiovert 200M inverted epifluorescence microscope using a 40x oil-immersion lens. Bodipy-fluorescein and carboxyfluorescein (495/20 excitation, 515 dichroic, 530/30 emission), Bodipy-TexasRed and SNARF-1 (560/20 excitation, 585 dichroic, 605/30 emission) and differential interference contrast (DIC) images (630/10 emission) were collected and analyzed using OpenLab software (Improvision).

Live-cell detection of azide 7 ligation using OCT-HDDA-CFDA

HEK 293T cells were transfected with the pcDNA3-E2p-CFP and pcDNA3-FLAG-LplA (Figure 8-5) or pcDNA3-mCherry-hLplA (Figure 8-7) plasmids using Lipofectamine 2000 (0.6 μg of E2p DNA and 0.2 μg of LplA DNA per well of a 48-well plate). 24 hours after transfection, 500 μM azide **7** was added to the cell growth media and cells were incubated at 37 °C for an additional 16–24 hours. Thereafter, azide **7** was removed from the cells by changing cell media three times every 20 minutes. The [3+2] cycloaddition reaction was initiated by addition of cell-loading media (5 μM OCT-HDDA-CFDA, 0.1% pluronic, 1% DMSO in DMEM with 10% FBS, 50 units/mL penicillin, and 50 μg/mL streptomycin) and incubation at 32 °C for 1–2 hours. Cells were washed and imaged as above except that images were analyzed using Slidebook software (Intelligent Imaging Innovations). The following filters were used to image CFP: 420/20 excitation, 450 dichroic, 475/40 emission. mCherry was imaged using the following filters: 560/20 excitation, 585 dichroic, 605/30 emission.

Fixed-cell immunostaining

HEK 293T or HeLa cells were fixed with 3.7% paraformaldehyde (PFA) for 5 minutes at 4 °C, followed by 10 minutes at room temperature. Excess PFA was washed three times, 5 minutes each wash, with DPBS and quenched with 100 mM glycine, pH 7.2 for 10 minutes at room temperature. After an additional 5-minute wash in DPBS, cells were permeabilized by incubation with 100% methanol for 6 minutes at –20 °C. Finally, cells were rinsed twice for 10 minutes with DPBS and blocked with 3% bovine serum albumin (BSA) in DPBS for 1 hour–overnight at room temperature or 4 °C with rocking. Antigen detection was performed by incubating cells with 5 μg/mL of the appropriate primary antibody for 1 hour–overnight at room temperature in a humidified chamber. After three 10-minute washes with 1% BSA and 0.03% Tween-20 in DPBS, cells were incubated for 1 hour with 6.7 μg/mL anti-rabbit or anti-mouse antibody conjugated to either Alexa Fluor 488 or Alexa Fluor 568 (Invitrogen). After washing, the cells were imaged in DPBS as above. Alexa Fluor 568 (560/20 excitation, 585 dichroic, 605/30 emission), Alexa Fluor 488 (495/20 excitation, 515 dichroic, 530/30 emission),

and differential interference contrast (DIC) images (630/10 emission) were collected and analyzed using OpenLab software (Improvision) or Slidebook software (Intelligent Imaging Innovations). The following primary antibodies were used: mouse monoclonal anti-FLAG®-M2 (Sigma, Cat# F3165) and rabbit polyclonal anti-lipoic acid (Calbiochem, Cat# 437695).

Detection of intracellular lipoic acid ligation to E2p by western blotting

Human embryonic kidney (HEK) 293T cells were transfected with the pcDNA3-E2p-CFP and pcDNA3-mCherry/FRB-e/hLplA plasmids using Lipofectamine 2000 (0.8 μg of each DNA per well of a 24-well plate). 24 hours after transfection, 500 μM R-(+)-lipoic acid was added to the cell growth media (DMEM with 10% FBS, 50 units/mL penicillin, and 50 μg/mL streptomycin). Lysates were generated 24 hours later by hypotonic lysis to minimize protease release, as described above for the detection of intracellular azide **7** ligation with biotin-alkyne. Cell extracts were analyzed by 12% SDS-PAGE followed by Western blotting with anti-lipoic antibody (Calbiochem, 1:1000 dilution) and anti-rabbit-peroxidase antibody conjugate (Pierce, 1:2000 dilution).

Detection of intracellular lipoic acid ligation to E2p in fixed cells by immunostaining

Human embryonic kidney (HEK) 293T cells were transfected with the pcDNA3-E2p-CFP and pcDNA3-mCherry-LplA plasmids using Lipofectamine 2000 (0.6 μg of E2p DNA and 0.2 μg of LplA DNA per well of a 48-well plate). 24 hours after transfection, 500 μM R-(+)-lipoic acid was added to the cell growth media (DMEM with 10% FBS, 50 units/mL penicillin, and 50 μg/mL streptomycin). Cells were fixed and stained 24 hours later as described above (fixed-cell immunostaining).

Fixed-cell detection of azide 7 ligation using Cy3-alkyne

HEK 293T cells were transfected with the pcDNA3-E2p-CFP and pcDNA3-FRB-LplA plasmids using Lipofectamine 2000 (0.4 μg of each DNA per well of a 48-well plate). From the moment of transfection, cells were grown in media supplemented with 750 μM azide **7**. 48 hours after transfection, azide **7** was removed from the cells by changing cell media three times every 20 minutes, and cells were fixed as described above. After blocking, azide derivatization via the [3+2] cycloaddition was performed by adding Cy3-alkyne to a final concentration of 50 μM in the presence of 1 mM $CuSO_4$, 2 mM tris(triazolyl)amine ligand (Sharpless ligand), and 4 mM tris(2-carboxyethyl)phosphine (TCEP) and incubating at room temperature for 16–24 hours. After three 10-minute washes with 1% BSA and 0.03% Tween-20 in DPBS, cells were imaged in DPBS as described above. Cy3 was imaged using the following filters: 560/20 excitation, 585 dichroic, 605/30 emission.

Live-cell detection of coumarin ligation

T-REx™-293 cells were transfected with the pcDNA3-E2p-CFP and pcDNA4-TO-mCherry-hLplA(W37I) plasmids using Lipofectamine 2000 (0.4 μg of each DNA per well of a 48-well plate). From the moment of transfection, cells were grown in media supplemented with 0.5 μg/mL tetracycline to induce protein expression. 48 hours after transfection, cell-loading media containing 500 μM of coumarin-acetate and 0.1% pluronic was added to the cells for 2 hours at 32 °C. Cells were washed for 1 hour and, thereafter, labeled cells were imaged as described above. The coumarin filter set was 405/20 excitation, 425 dichroic, 460/40 emission.

References

1. Keppler,A. *et al.* A general method for the covalent labeling of fusion proteins with small molecules in vivo. *Nat. Biotechnol.* **21**, 86-89 (2003).

2. Los,G.V. & Wood,K. The HaloTag: a novel technology for cell imaging and protein analysis. *Methods Mol. Biol.* **356**, 195-208 (2007).

3. Chen,I., Howarth,M., Lin,W. & Ting,A.Y. Site-specific labeling of cell surface proteins with biophysical probes using biotin ligase. *Nat. Methods* **2**, 99-104 (2005).

4. Zhou,Z. *et al.* Genetically encoded short peptide tags for orthogonal protein labeling by Sfp and AcpS phosphopantetheinyl transferases. *ACS Chem. Biol.* **2**, 337-346 (2007).

5. Keppler,A., Arrivoli,C., Sironi,L. & Ellenberg,J. Fluorophores for live cell imaging of AGT fusion proteins across the visible spectrum. *Biotechniques* **41**, 167-5 (2006).

6. Zlokarnik,G. *et al.* Quantitation of transcription and clonal selection of single living cells with beta-lactamase as reporter. *Science* **279**, 84-88 (1998).

7. Ali,S.T. & Guest,J.R. Isolation and characterization of lipoylated and unlipoylated domains of the E2p subunit of the pyruvate dehydrogenase complex of Escherichia coli. *Biochem. J.* **271**, 139-145 (1990).

8. Chen,I. Site-specific labeling of cellular proteins with unnatural substrates of biotin ligases. 2007.

9. Rengelshausen,J. *et al.* Inhibition of the human organic anion transporter 1 by the caffeine metabolite 1-methylxanthine. *Biochem. Biophys. Res. Commun.* **320**, 90-94 (2004).

10. Dumitriu,I.E. *et al.* UV irradiation inhibits ABC transporters via generation of ADP-ribose by concerted action of poly(ADP-ribose) polymerase-1 and glycohydrolase. *Cell Death. Differ.* **11**, 314-320 (2004).

11. Calloway,N.T. *et al.* Optimized fluorescent trimethoprim derivatives for in vivo protein labeling. *Chembiochem.* **8**, 767-774 (2007).

12. Farinas,J. & Verkman,A.S. Receptor-mediated targeting of fluorescent probes in living cells. *J. Biol. Chem.* **274**, 7603-7606 (1999).

13. Griffin,B.A., Adams,S.R. & Tsien,R.Y. Specific covalent labeling of recombinant protein molecules inside live cells. *Science* **281**, 269-272 (1998).

14. Miller,L.W., Cai,Y., Sheetz,M.P. & Cornish,V.W. In vivo protein labeling with trimethoprim conjugates: a flexible chemical tag. *Nat. Methods* **2**, 255-257 (2005).

15. Keppler,A., Pick,H., Arrivoli,C., Vogel,H. & Johnsson,K. Labeling of fusion proteins with synthetic fluorophores in live cells. *Proc. Natl. Acad. Sci. U. S. A* **101**, 9955-9959 (2004).

16. Pedelacq,J.D., Cabantous,S., Tran,T., Terwilliger,T.C. & Waldo,G.S. Engineering and characterization of a superfolder green fluorescent protein. *Nat. Biotechnol.* **24**, 79-88 (2006).

17. Yao,F. *et al.* Tetracycline repressor, tetR, rather than the tetR-mammalian cell transcription factor fusion derivatives, regulates inducible gene expression in mammalian cells. *Hum. Gene Ther.* **9**, 1939-1950 (1998).

18. Baskin,J.M. *et al.* Copper-free click chemistry for dynamic in vivo imaging. *Proc. Natl. Acad. Sci. U. S. A* **104**, 16793-16797 (2007).

19. Rust,M.J., Bates,M. & Zhuang,X. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nat. Methods* **3**, 793-795 (2006).

20. Bates,M., Huang,B., Dempsey,G.T. & Zhuang,X. Multicolor super-resolution imaging with photo-switchable fluorescent probes. *Science* **317**, 1749-1753 (2007).

21. Huang,B., Wang,W., Bates,M. & Zhuang,X. Three-dimensional super-resolution imaging by stochastic optical reconstruction microscopy. *Science* **319**, 810-813 (2008).

22. Agard,N.J., Baskin,J.M., Prescher,J.A., Lo,A. & Bertozzi,C.R. A comparative study of bioorthogonal reactions with azides. *ACS Chem. Biol.* **1**, 644-648 (2006).

Marta Fernández Suárez

Phone: (857) 998 0592 • e-mail: martafs@alum.mit.edu

EDUCATION

2003-2008 Ph.D., Chemistry, **Massachusetts Institute of Technology** (Cambridge, MA).

2000-2001 M.S., Chemical Engineering, **Institut Químic de Sarrià,** Universitat Ramon Llull (Barcelona, Spain).

1995-1999 B.S., Dual degree Chemical and Industrial Engineering, **Institut Químic de Sarrià,** Universitat Ramon Llull (Barcelona, Spain).

RESEARCH EXPERIENCE

2003-2008 Graduate Student, **Massachusetts Institute of Technology**, Department of Chemistry. Supervisor: Alice Y. Ting. Developed reporters for site-specific protein labeling and for detecting protein-protein interactions in live cells. Expertise acquired in organic chemistry, assay development, mammalian cell culture, fluorescence imaging, protein and enzyme biochemistry, molecular cloning, and RNA biochemistry. Supervised and mentored two undergraduate and two graduate students.

2000-2001 Research Assistant, **Massachusetts Institute of Technology**, Health Sciences & Technology Division, Biomedical Engineering Center. Supervisor: Elazer R. Edelman. Designed a bioreactor for subjecting cells to pulsatile flow *in vitro* and studied the frequency response of various cell types. Expertise acquired in surface chemistry, mammalian cell culture, simple electronic circuits, and construction of devices to exert fluid-mechanical forces on cultured cells.

1999 Research Intern at **BASF** AG (Ludwigshafen, Germany) in the Caprolactam production plant. Designed new distillation columns to optimize production efficiency. Columns were implemented one year later.

1997 Research Intern at **Asturiana de Zinc** S.A. (Asturias, Spain) in the R&D department. Researched electrolysis as method for purifying Zinc metal. Examined the role of pH, and found the optimum pH for the process. Optimum pH was implemented by the end of the internship.

FELLOWSHIPS AND AWARDS

2008 **Massachusetts General Hospital – Massachusetts Institute of Technology** Career Development Postdoctoral Fellowship in Translational Research to support two years of post-doctoral research.

2008 Grant from the **MIT International Development Initiative** to perform summer work in Honduras towards building a start-up to provide distributed biogas energy to rural farmers.

2008 People Helping People Award from the **MIT Federal Credit Union**.

2008 Semi-finalist of the **MIT 100K Business Plan Contest** for a start-up to maximize dissemination of distributed biogas energy for the rural population at the base of the pyramid.

2007 **Lewis Paul Chapin** Graduate Fellowship.

2007 Winner of the **Massachusetts Institute of Technology** IDEAS competition for a low-cost wind-powered water pump for developing countries.

2006 Travel grant awardee from the **National Institutes of Health** to attend the Frontiers in Live Cell Imaging conference (Bethesda, MD).

2003-2005 Graduate fellowship from the **La Caixa Foundation** (Spain).

2004 Winner of the **Massachusetts Institute of Technology** IDEAS competition for the development of a novel solution to water treatment in developing countries.

1994 National prize Premio Extraordinario de Bachillerato awarded by the **Spanish Government** for excellence in high school.

PUBLICATIONS

Fernández-Suárez, M., Ting, A.Y. "Fluorescent probes for super-resolution imaging in living cells." *Nature Reviews in Molecular and Cellular Biology*, invited review, manuscript in preparation.

Fernández-Suárez, M., Chen, S., Ting, A.Y. "Detecting cellular protein-protein interactions by enzymatic transfer of biotin to a re-engineered acceptor peptide." Submitted.

Fernández-Suárez, M., Baruah, H., Martínez-Hernández, L., Xie, K., Baskin, J., Bertozzi, C.R., Ting, A.Y. "Redirecting lipoic acid ligase for cell surface protein labeling with small-molecule probes." *Nature Biotechnology* **2007**, *25*, 1483-1487.

Balcells, M., Fernández-Suárez, M., Vázquez, M., Edelman, E.R. "Cells in fluidic environments are sensitive to flow frequency." *Journal of Cellular Physiology* **2005**, *204*, 329-335.

PATENTS

Fernández-Suárez, M., Baruah, H., Choi, Y., Ting, A.Y. "Methods and compositions for polypeptide and protein labeling using lipoic acid ligases" U.S. patent application no. 11/983605 (filed November 9, 2007).

INVITED TALKS

Apr. 2008 MIT Biological Chemistry division seminar series (Cambridge, MA).

Aug. 2007 National Meeting of the American Chemical Society (Boston, MA).

Dec. 2006 MIT Departments of Chemistry and Biology supergroup meeting (Cambridge, MA).

Nov. 2006 Harvard University Departments of Chemistry and Chemical Biology, Single Molecule Discussion Club seminar series (Cambridge, MA).

Oct. 2006 MIT Biological Engineering division seminar series (Cambridge, MA).

Jul. 2006 International Symposium on Optical Analysis of Biomolecular Machines (Berlin, Germany).

Dec. 2004 Broad Institute chemical biology meeting (Cambridge, MA).

POSTER PRESENTATIONS

Jul. 2007 21st Symposium of the Protein Society (Boston, MA). *Title: Re-directing lipoic acid ligase for site-specific labeling of receptors in living cells.*

Apr. 2006 Frontiers in Live Cell Imaging conference at the National Institutes of Health (Bethesda, MD). *Title: Site-specific protein labeling in living cells using lipoic acid ligase.*

TEACHING EXPERIENCE

2007 Teaching Assistant, **Massachusetts Institute of Technology**, Edgerton
2006 Center. 11-472 D-Lab Development. Instructor: Amy B. Smith.
2004 Advised and mentored teams of MIT undergraduate students working on International Development projects in Honduras. Class focuses on designing appropriate technologies to be implemented in Honduras at the end of the semester.

2003 Teaching assistant, **Massachusetts Institute of Technology**, Department of Chemistry. 5.12 Organic Chemistry I. Instructor: Prof. Barbara Imperiali. Prepared and delivered recitations to a group of twelve undergraduate students. Graded homework and exams. Class webmaster.

2001 Teaching assistant, **Maria L. Baldwin Public School** (Cambridge, MA) through the MIT LINKS program. Science lab for K-8. Assisted the teacher and prepared and delivered lab practice experiments. Focus was on supporting non-native English speakers of Hispanic origin